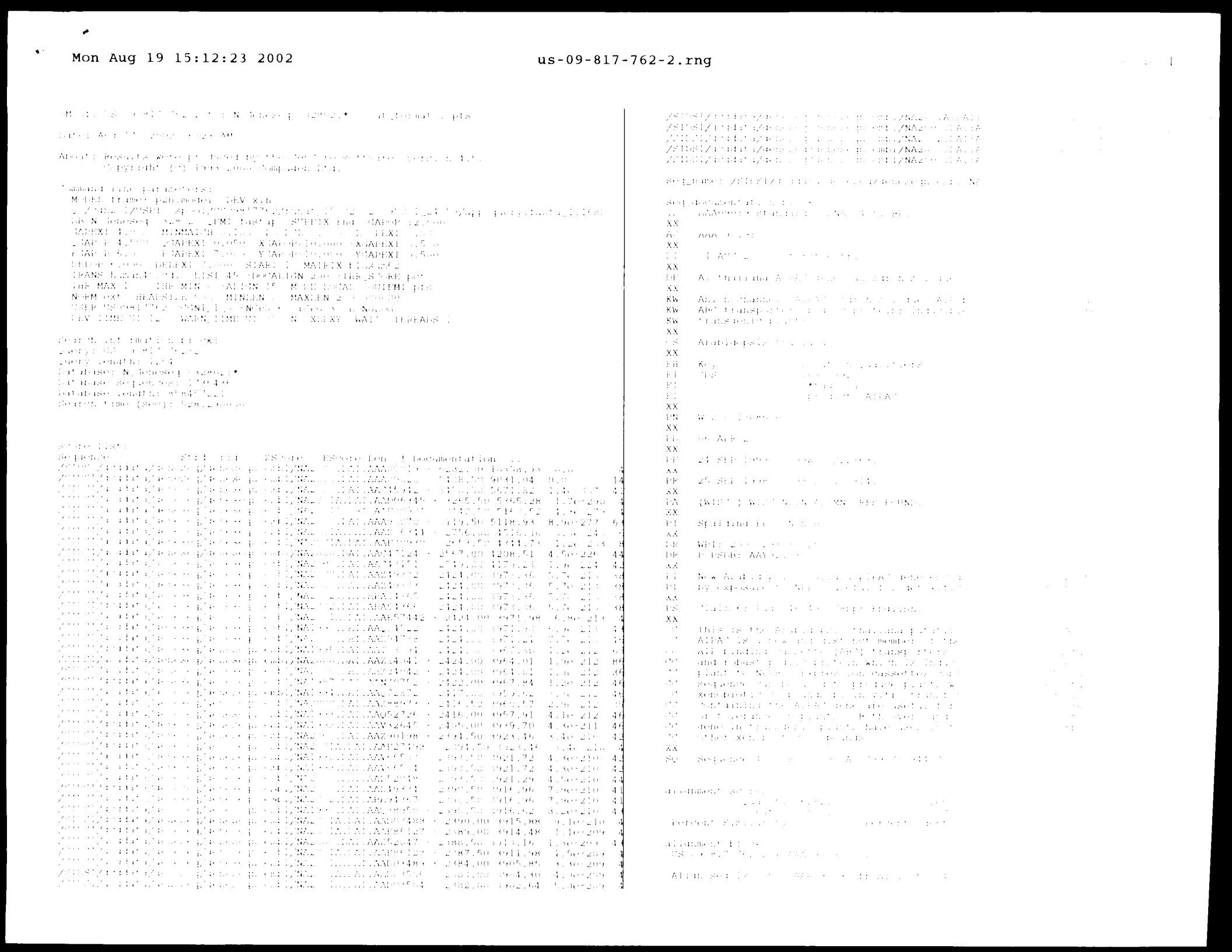
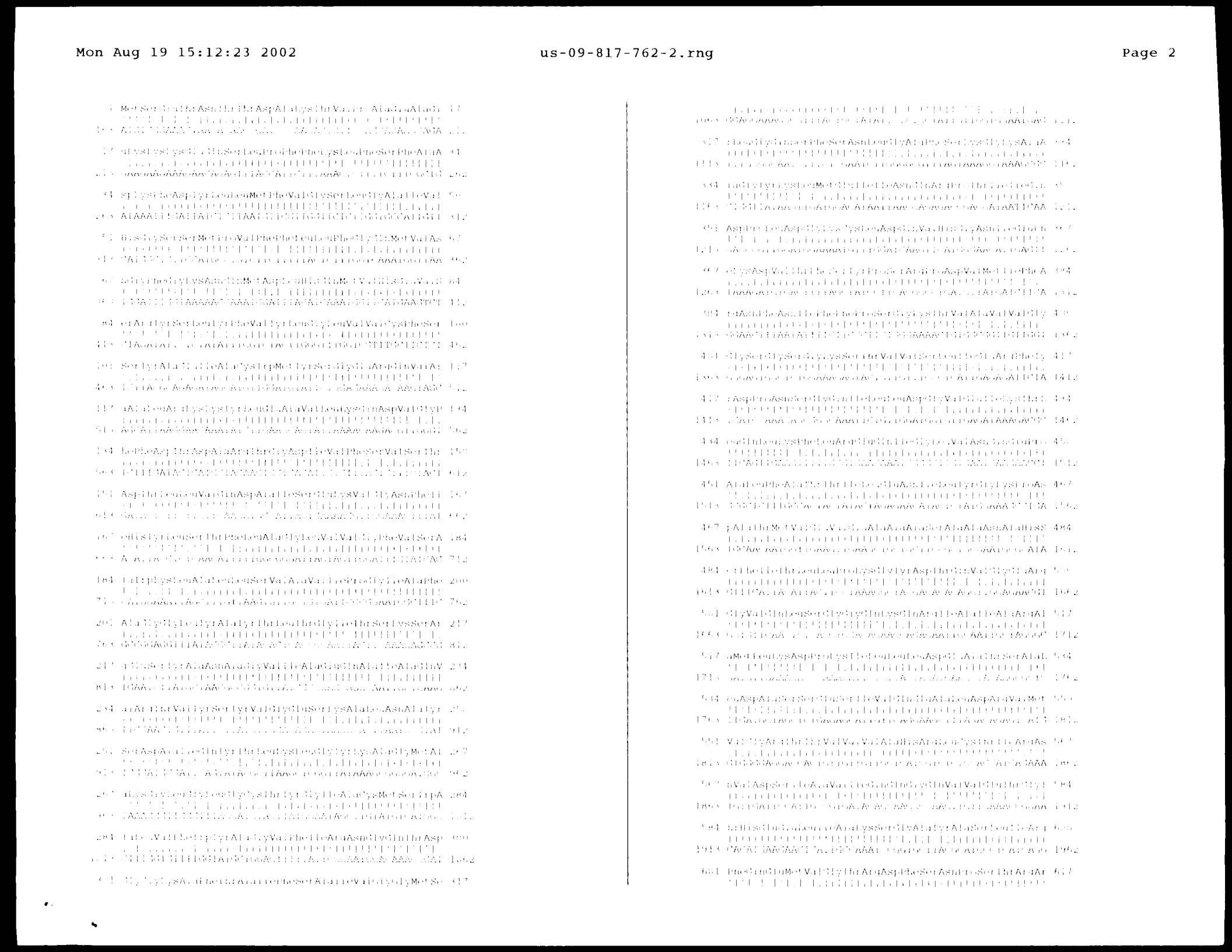
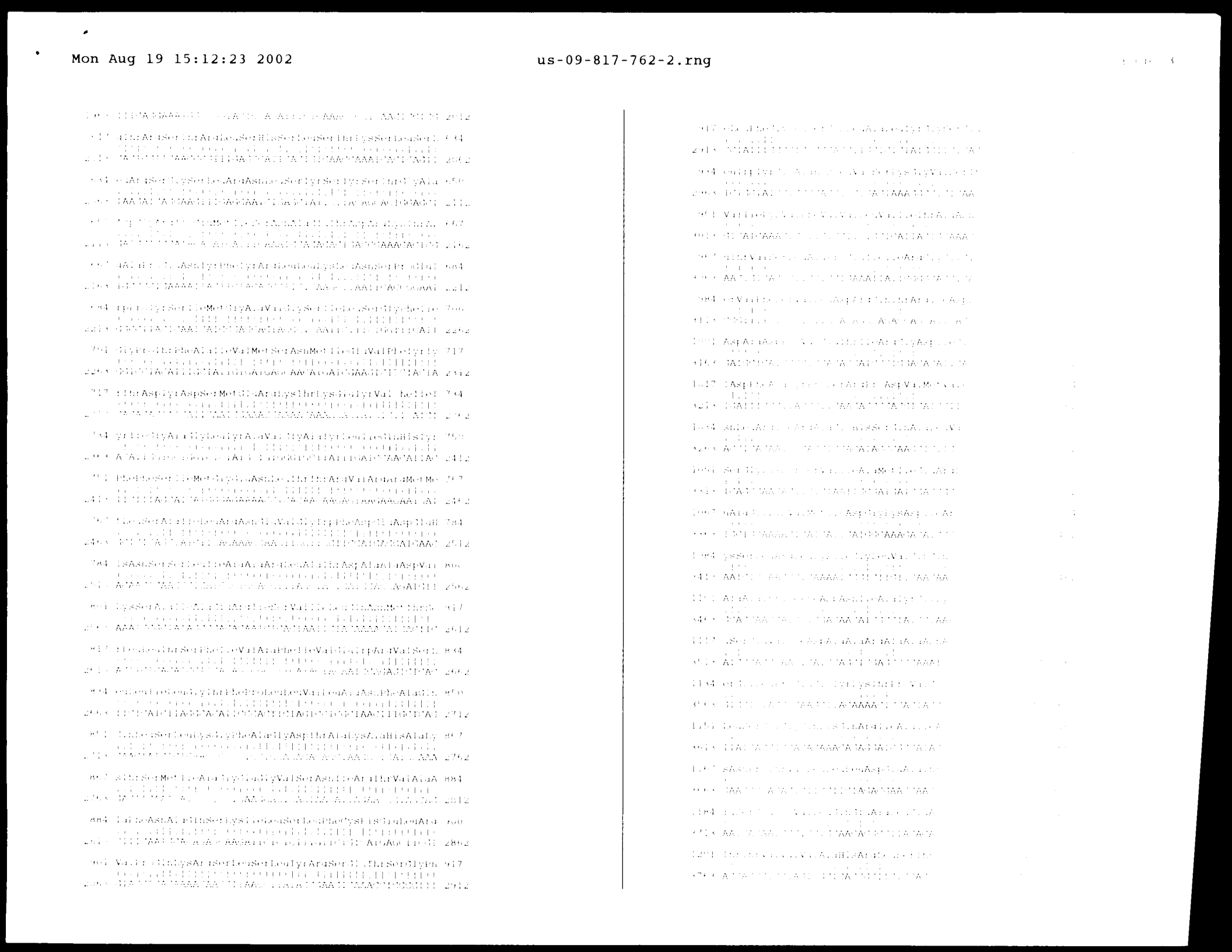
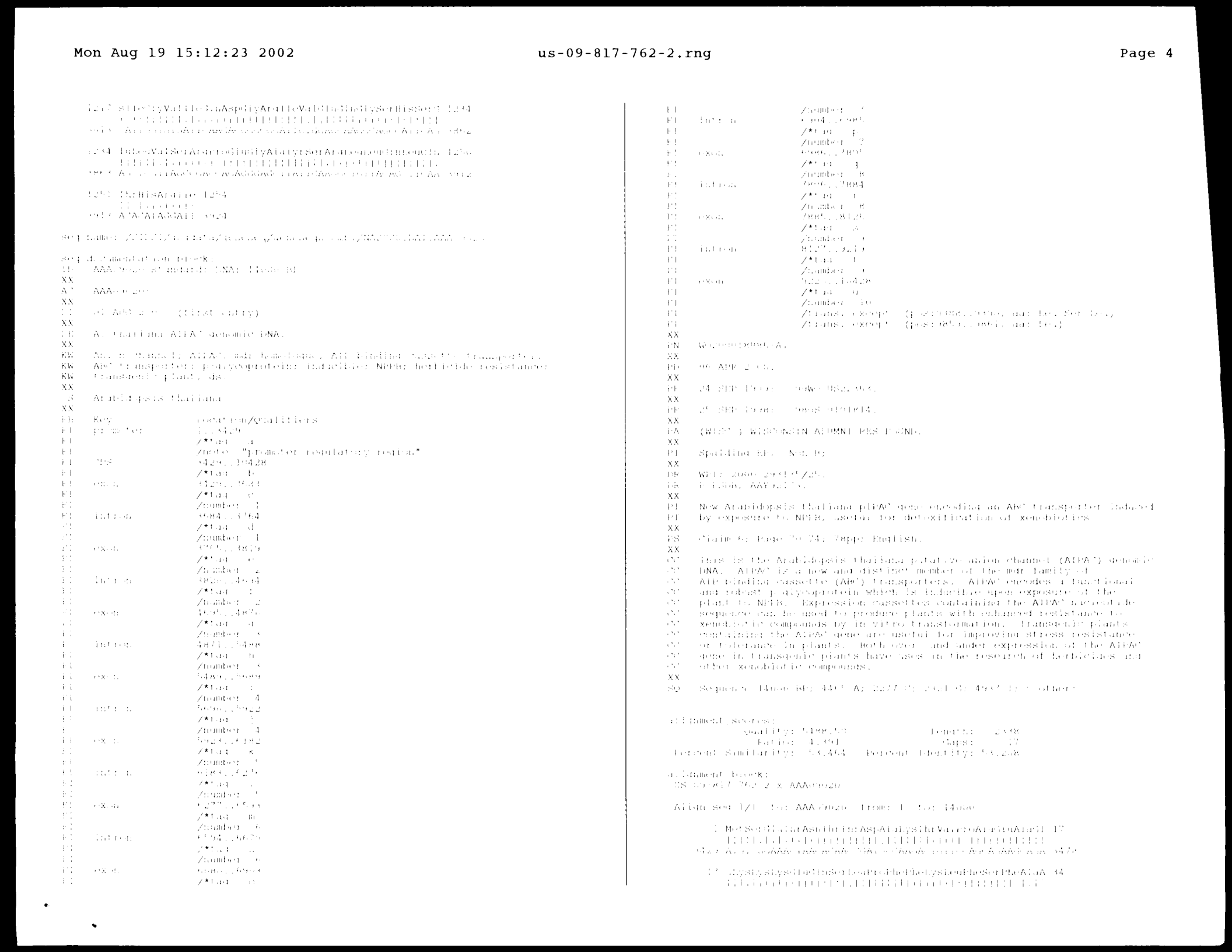
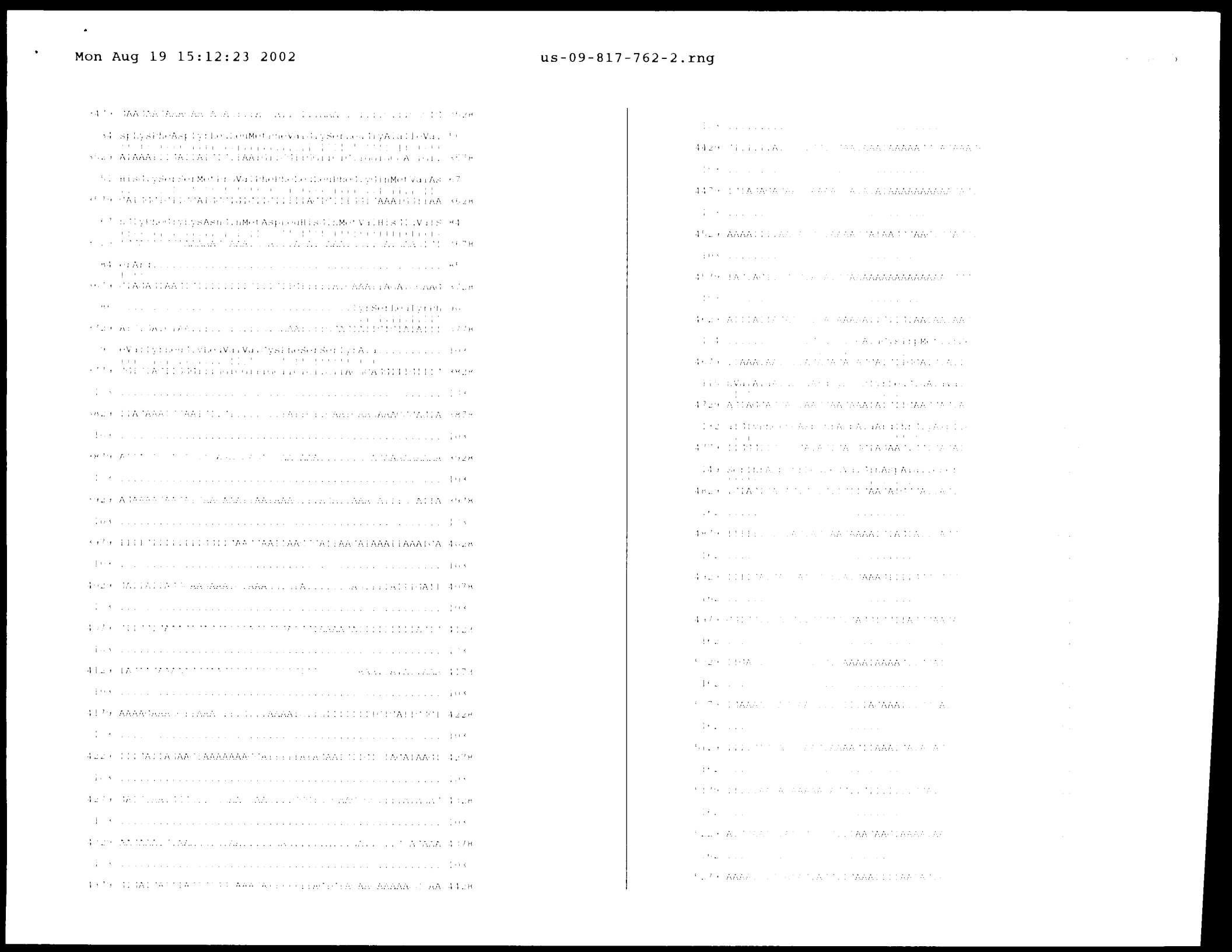
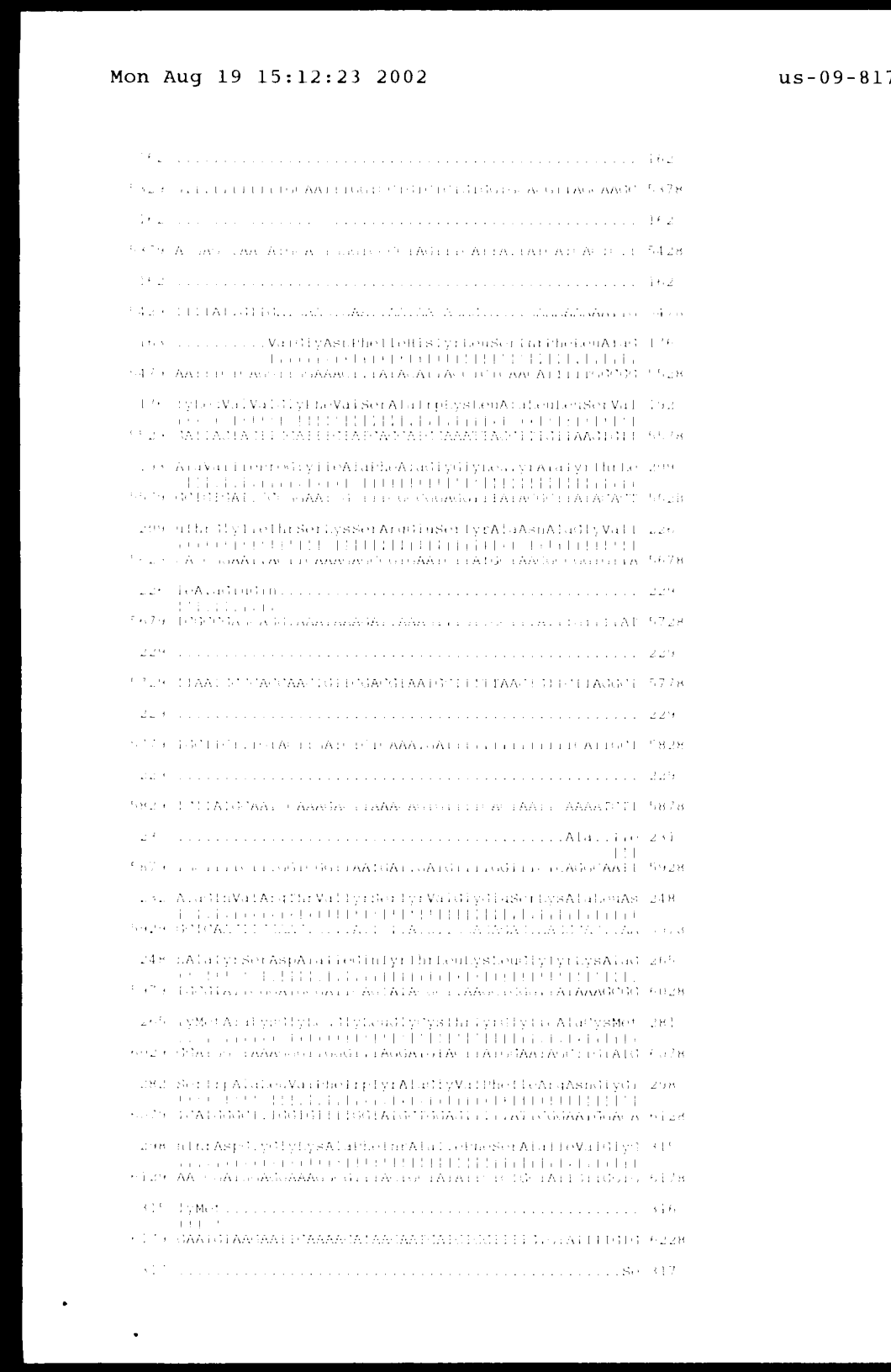


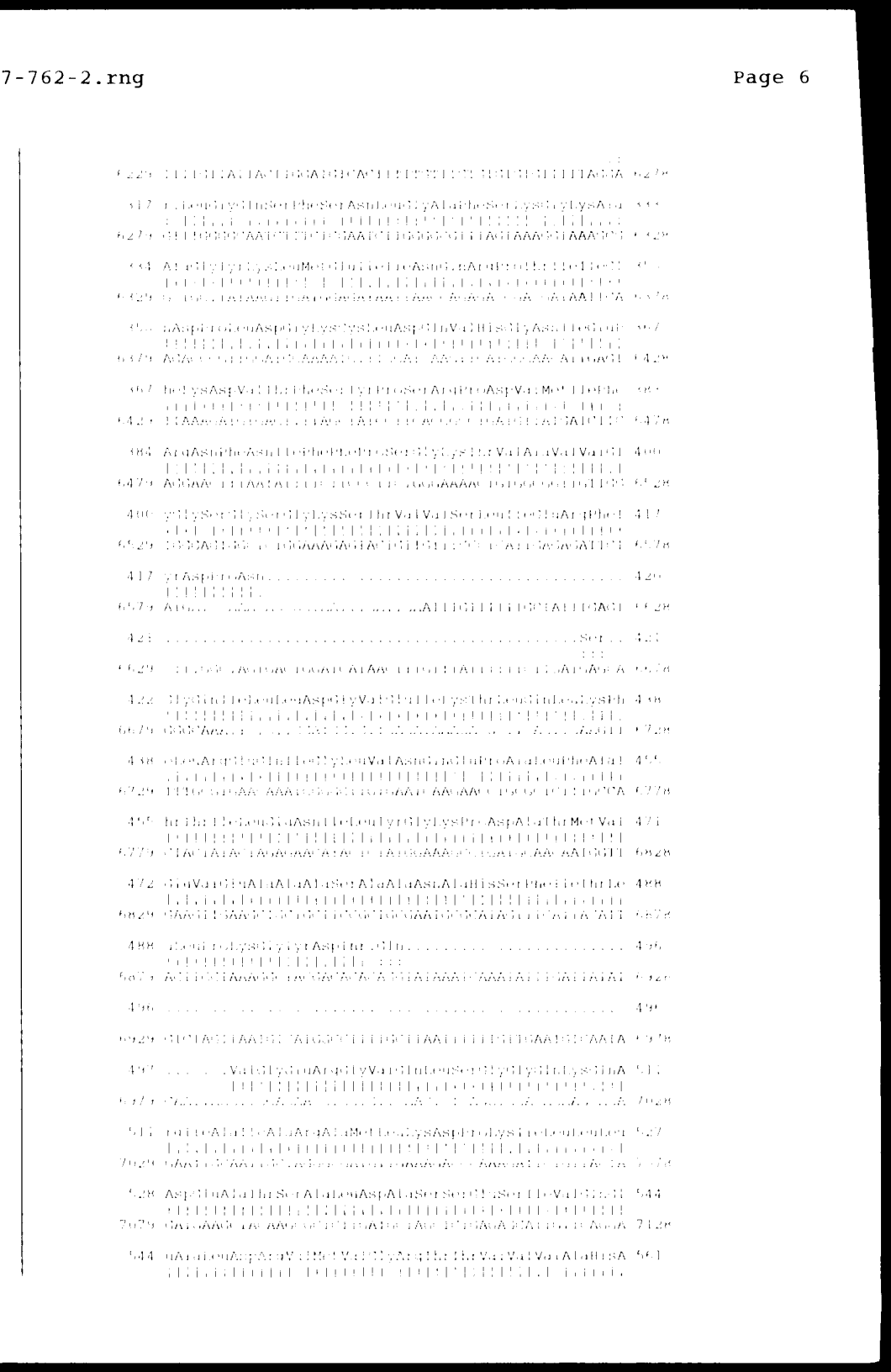


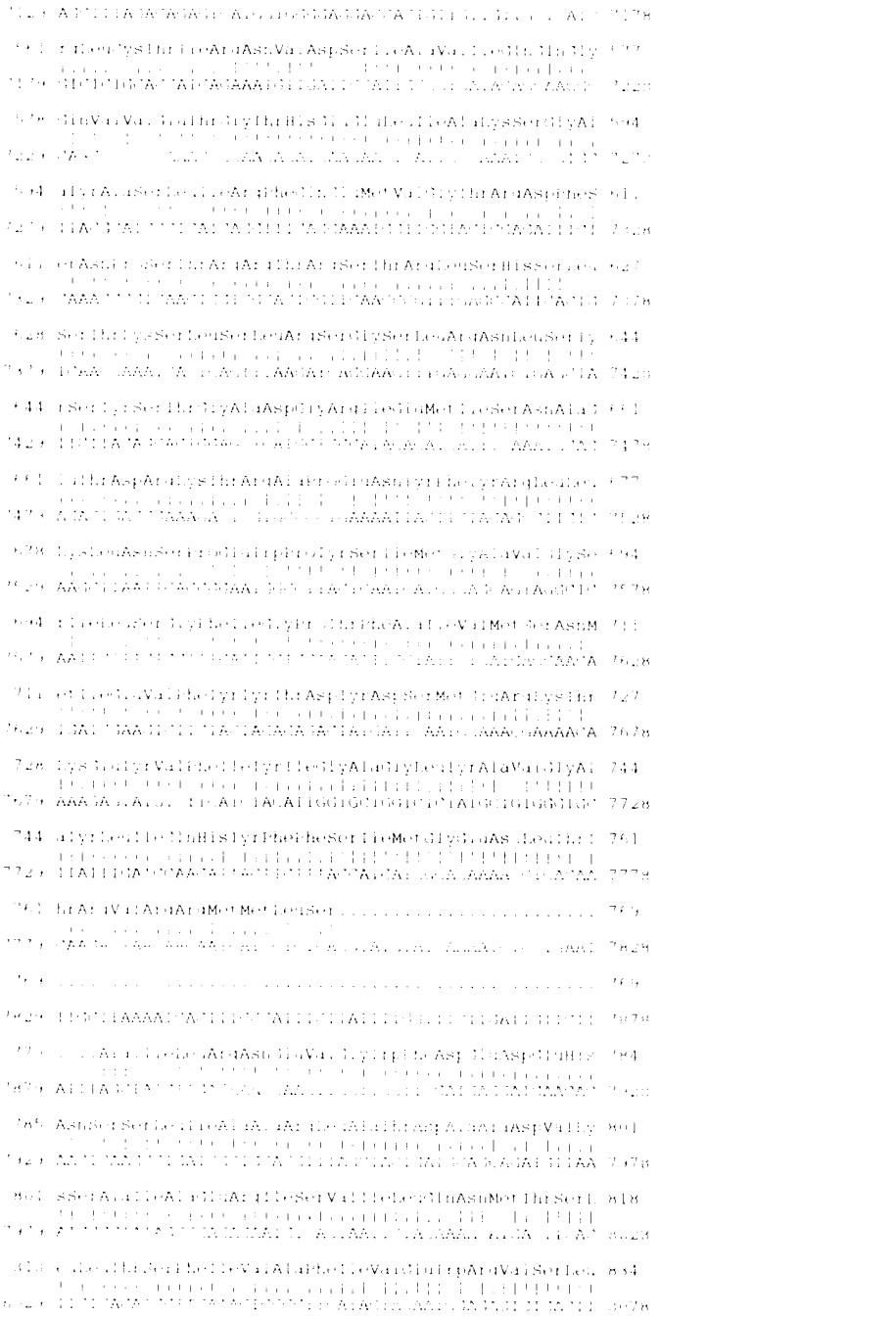


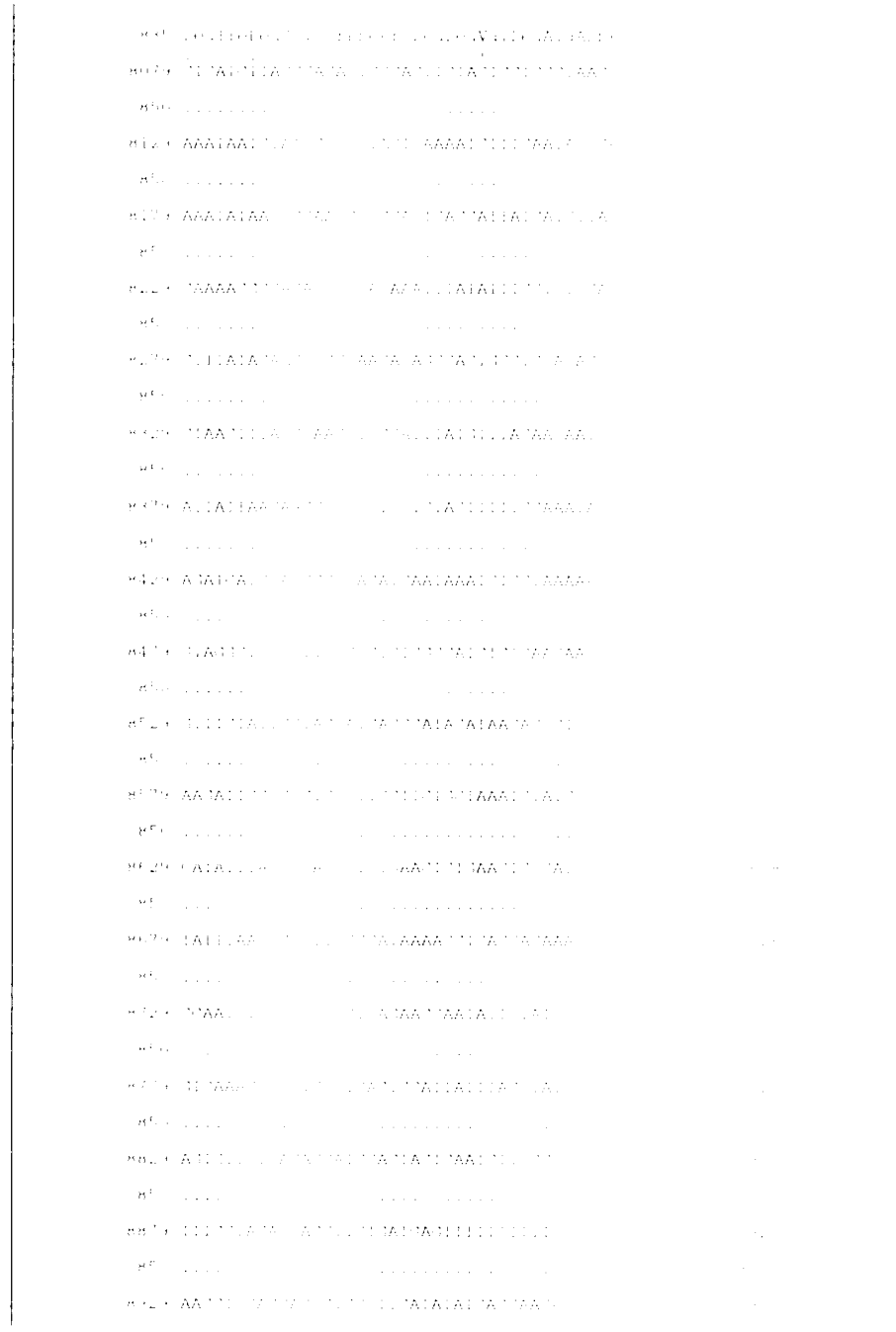


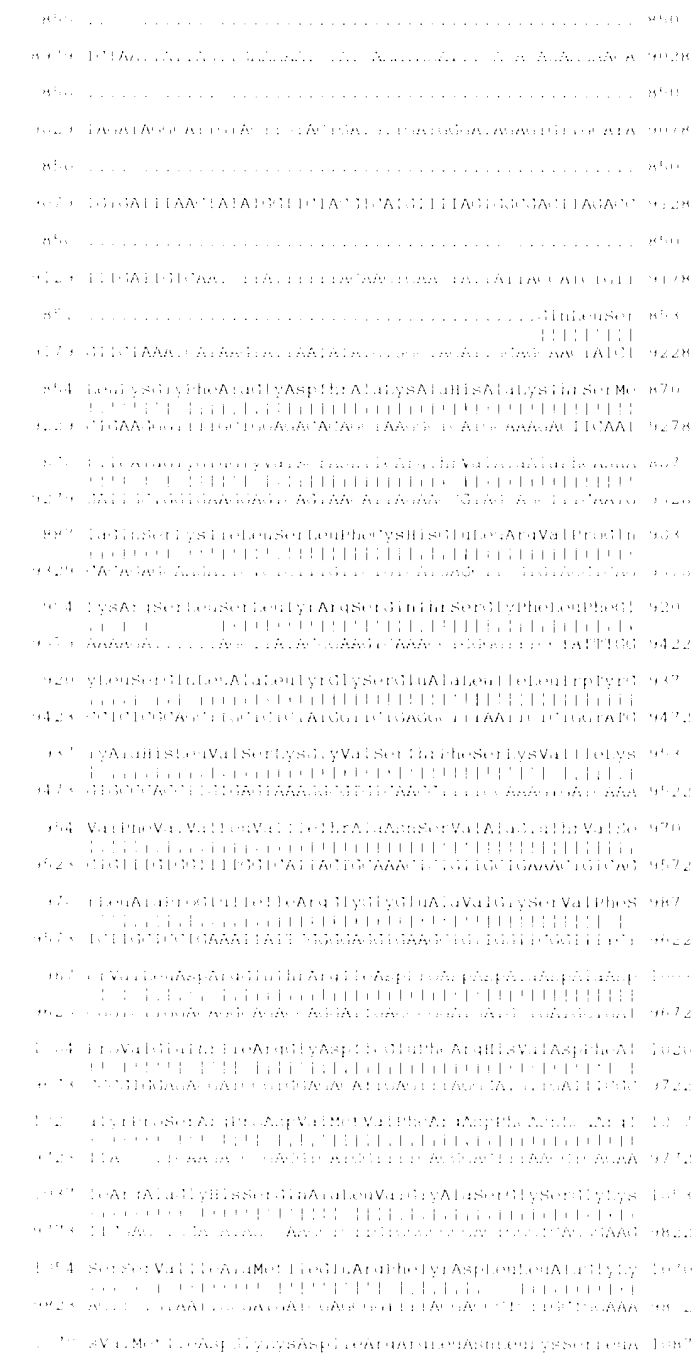


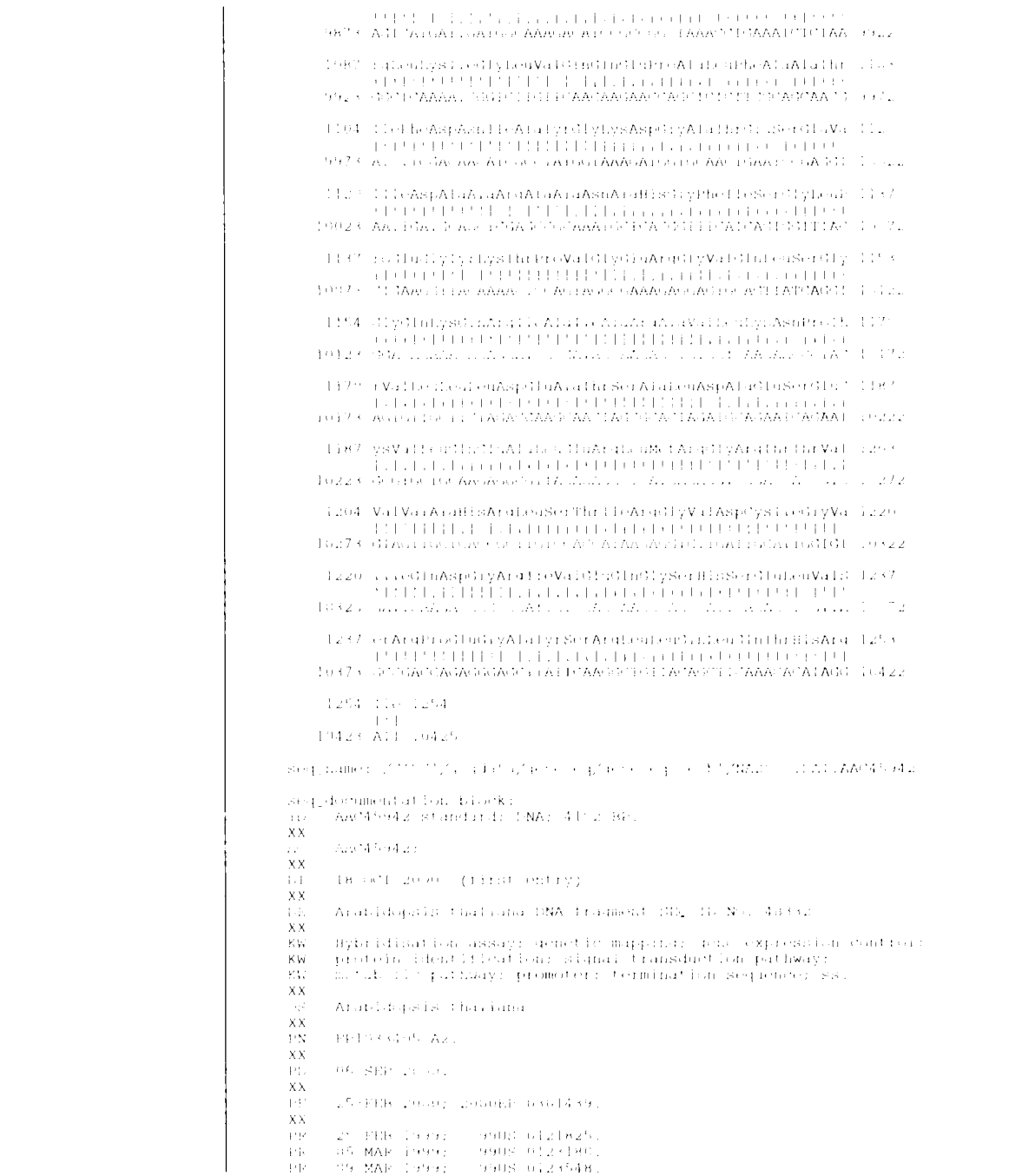


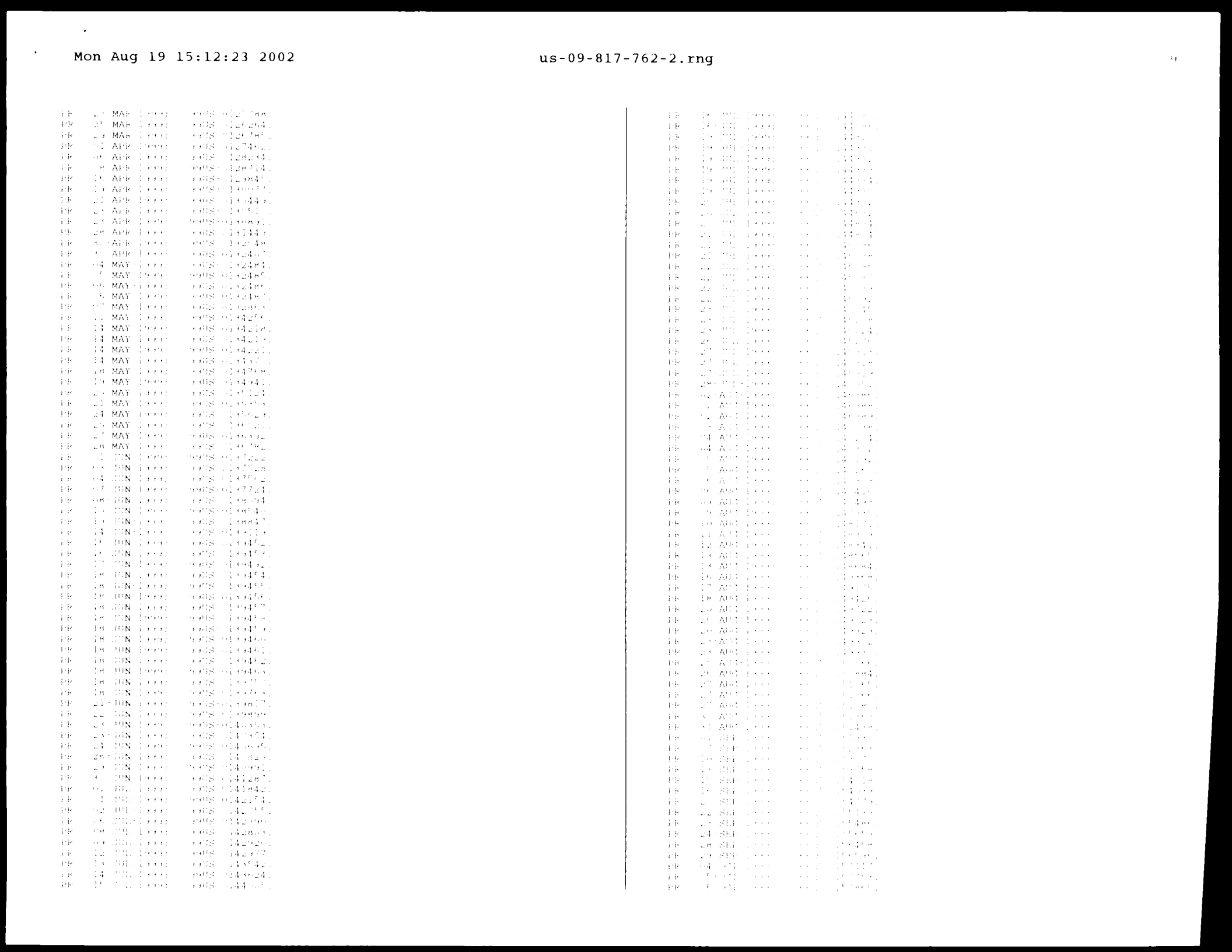


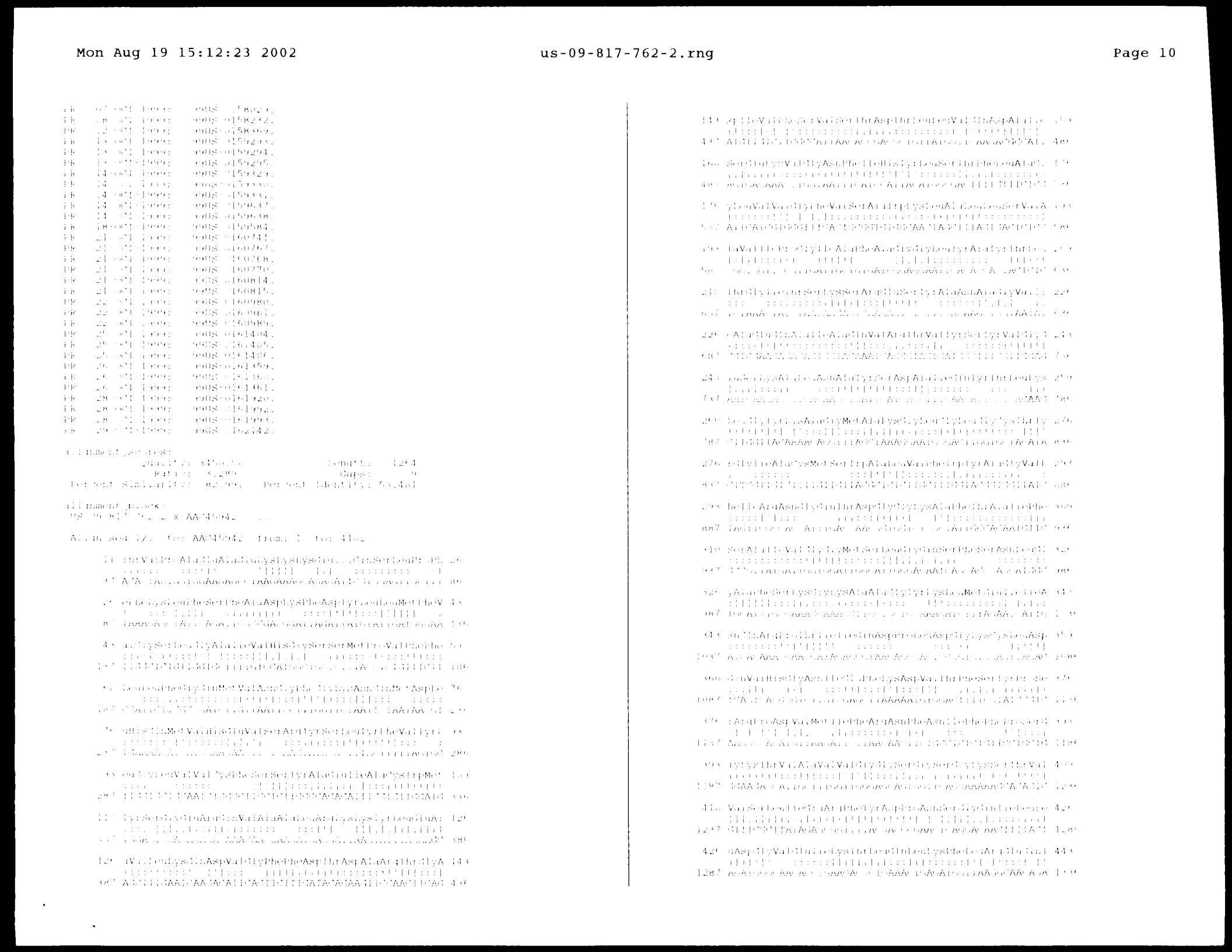


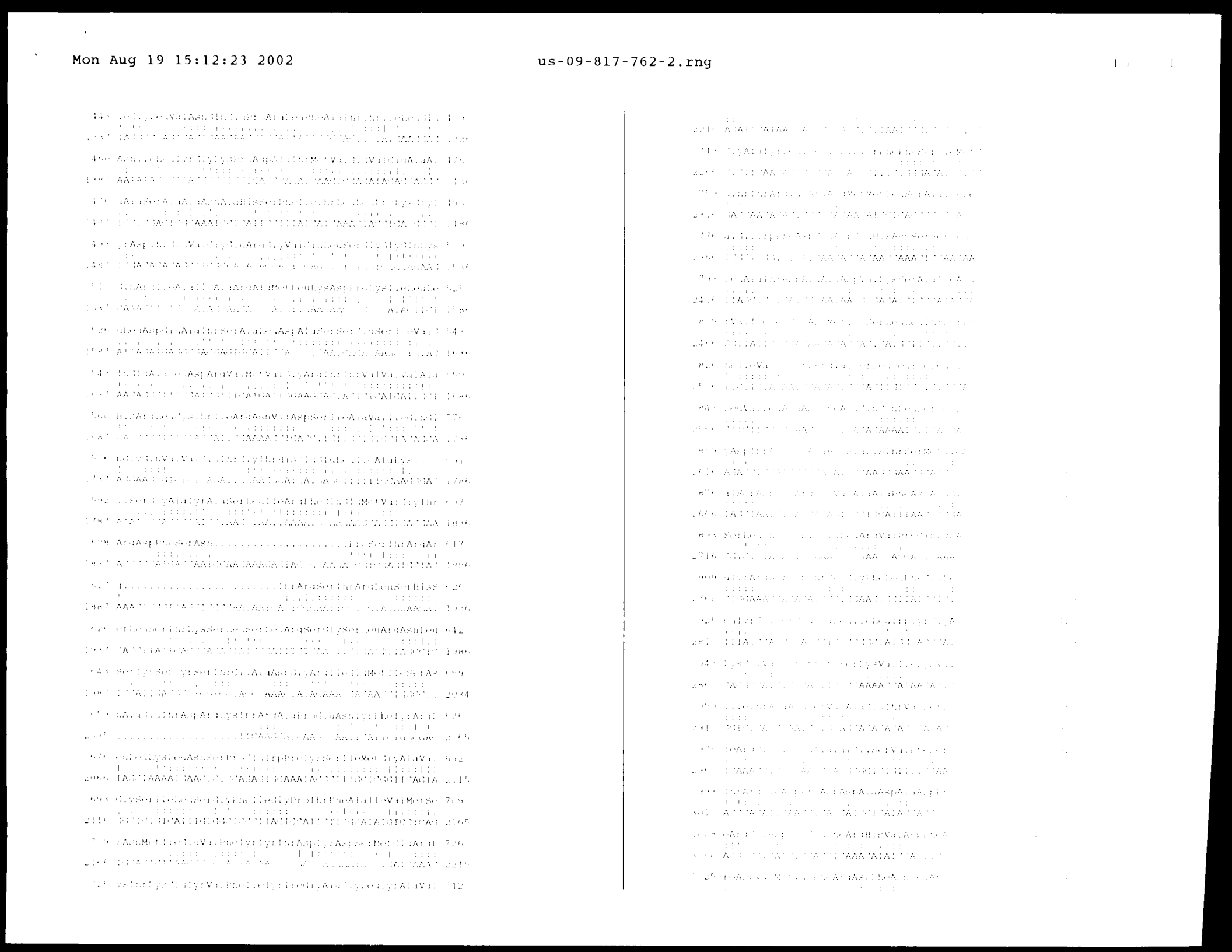


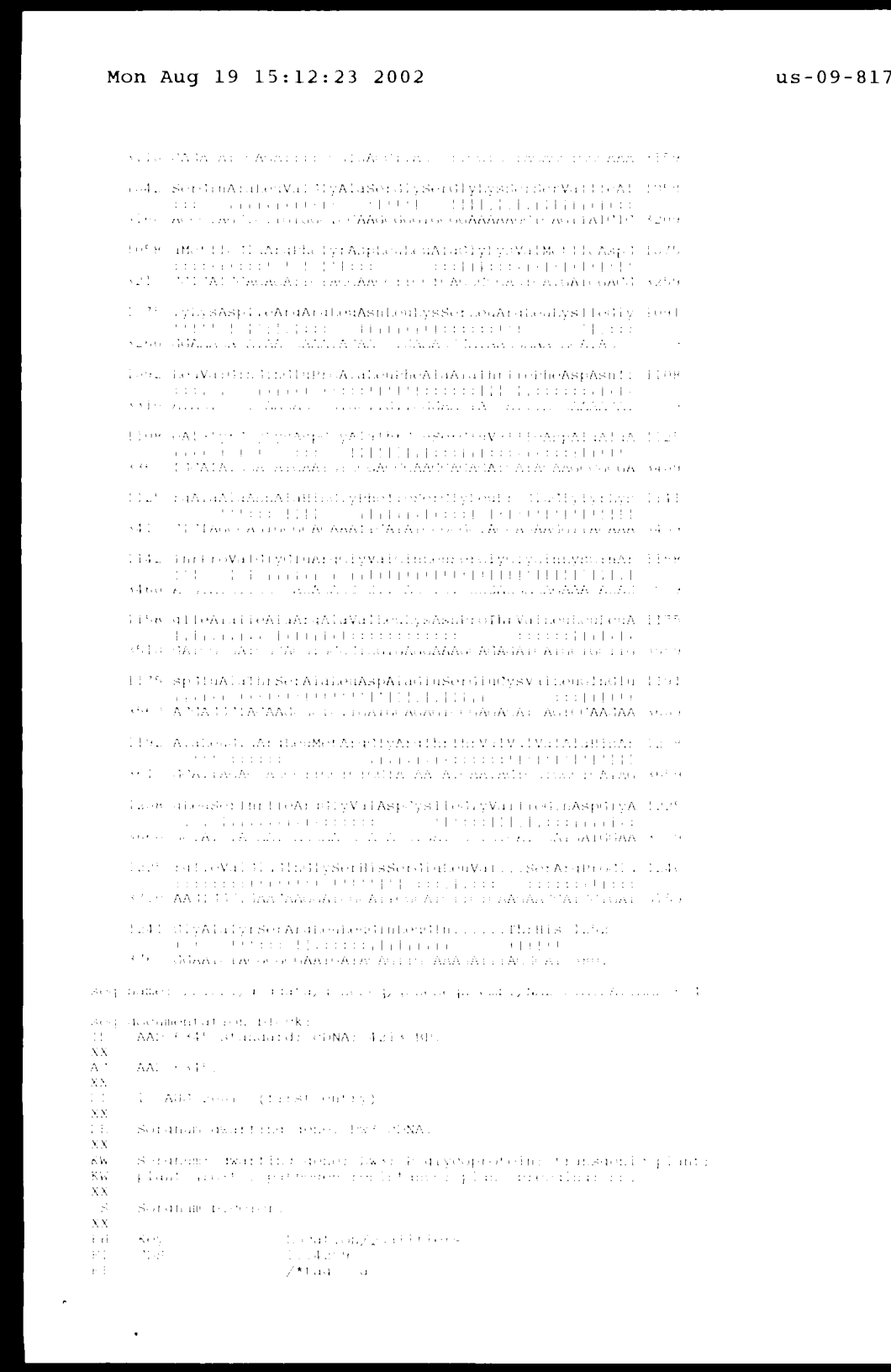


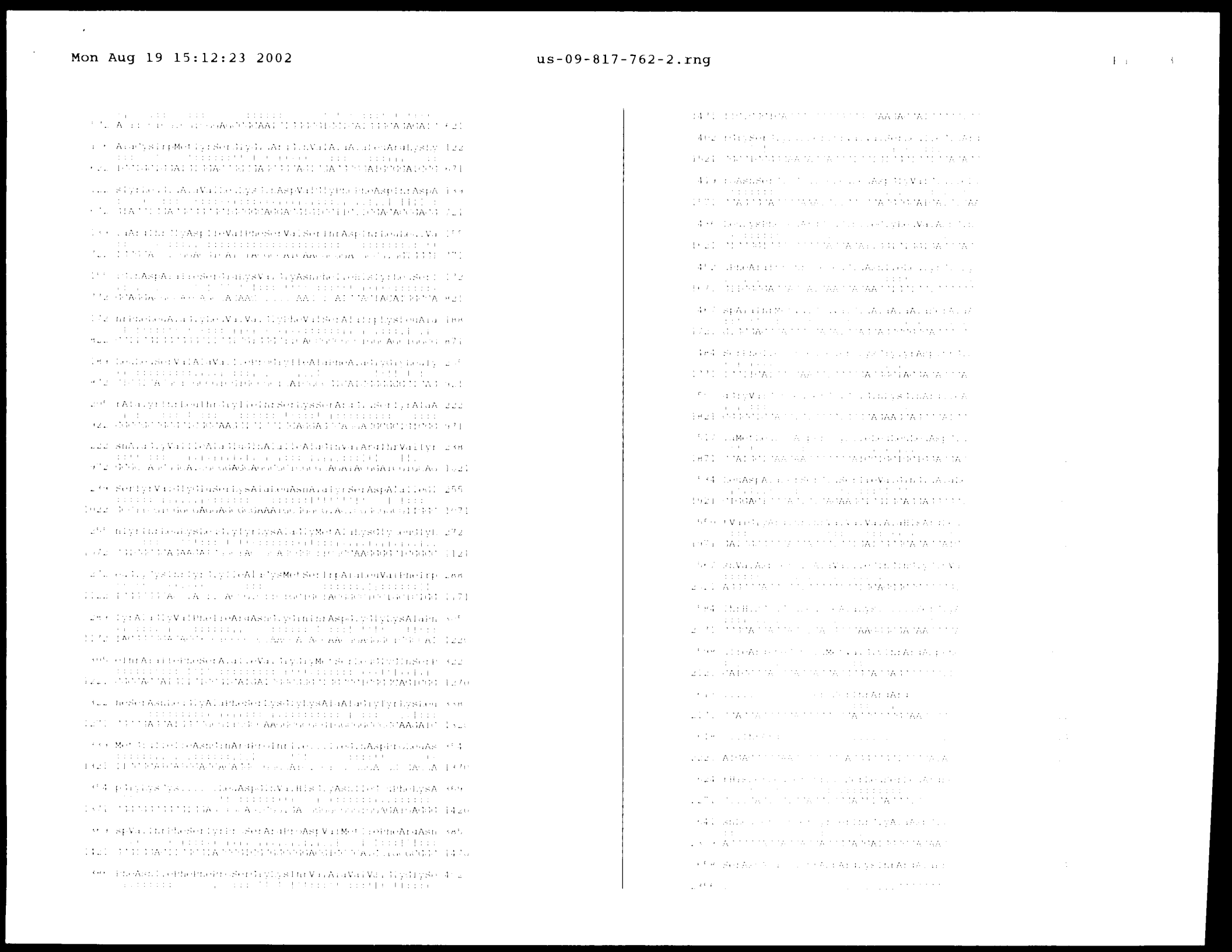


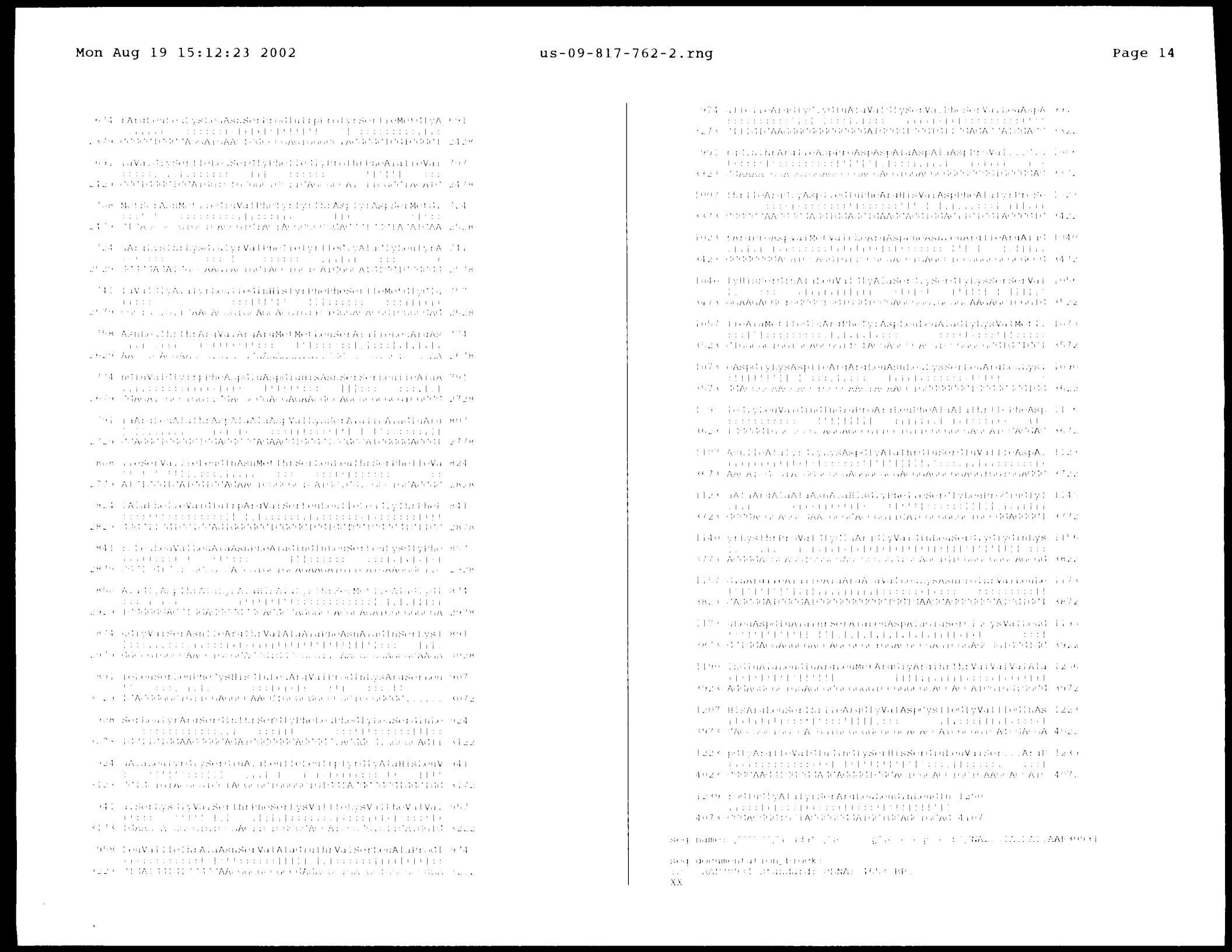






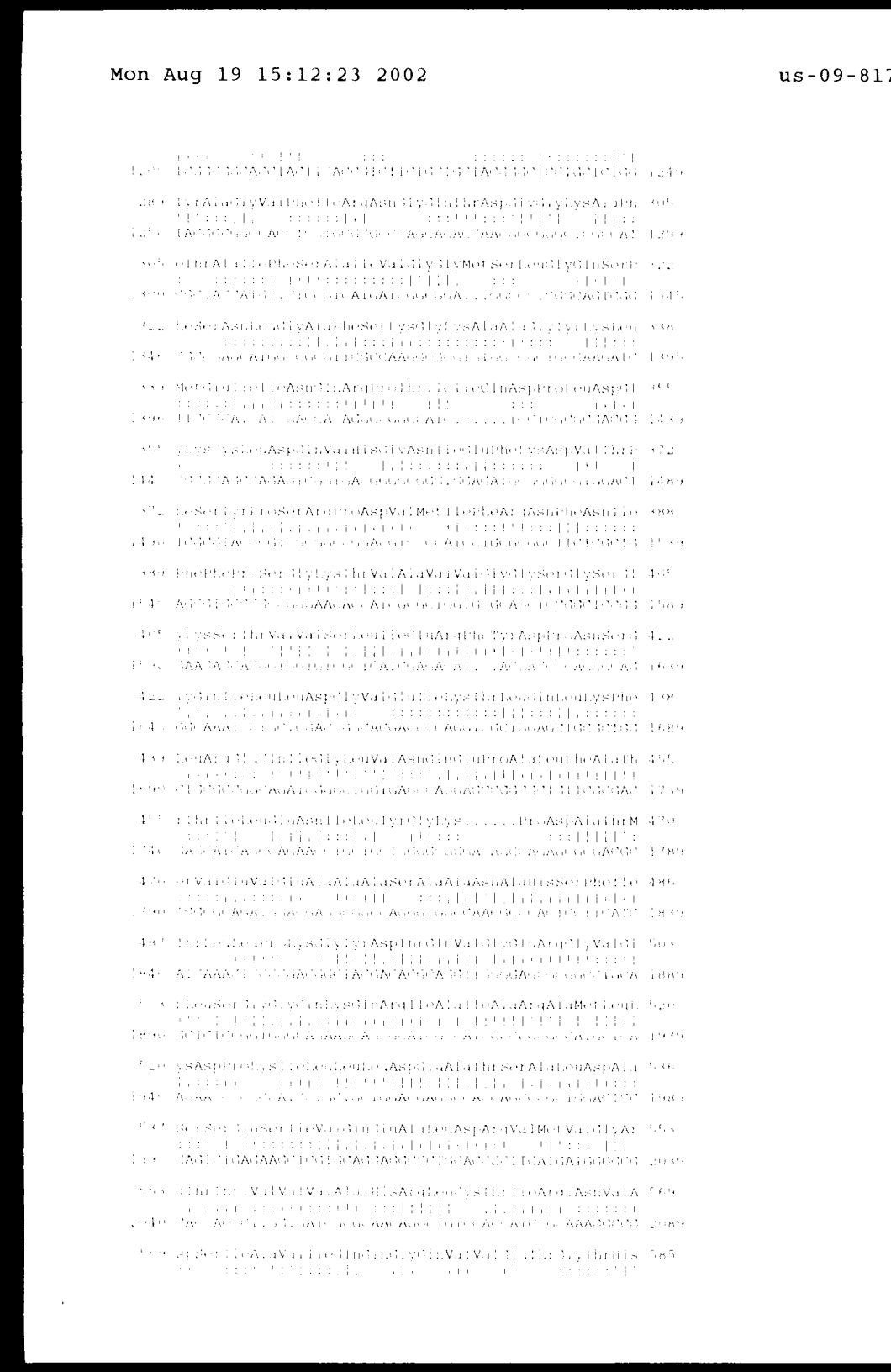


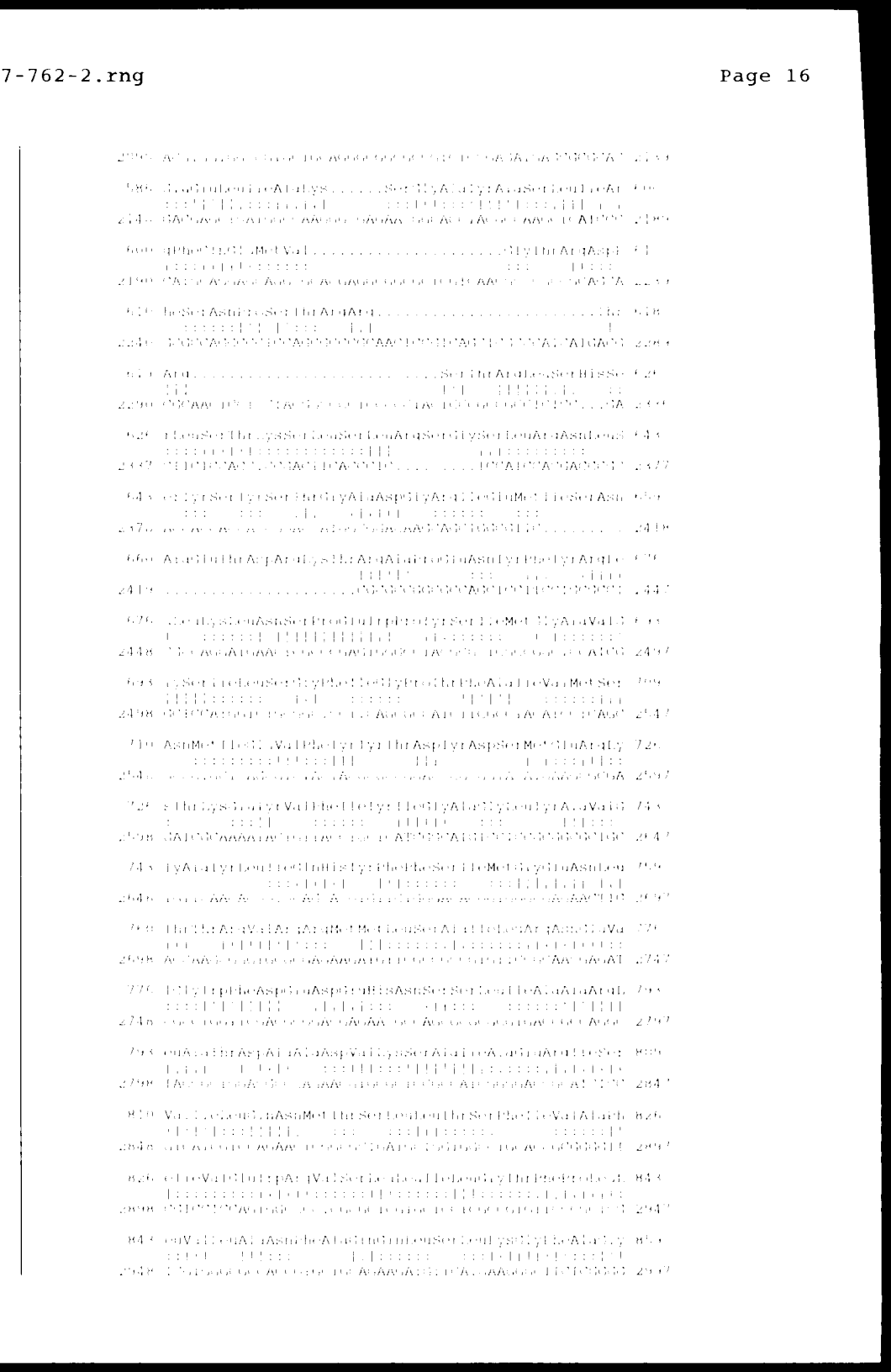


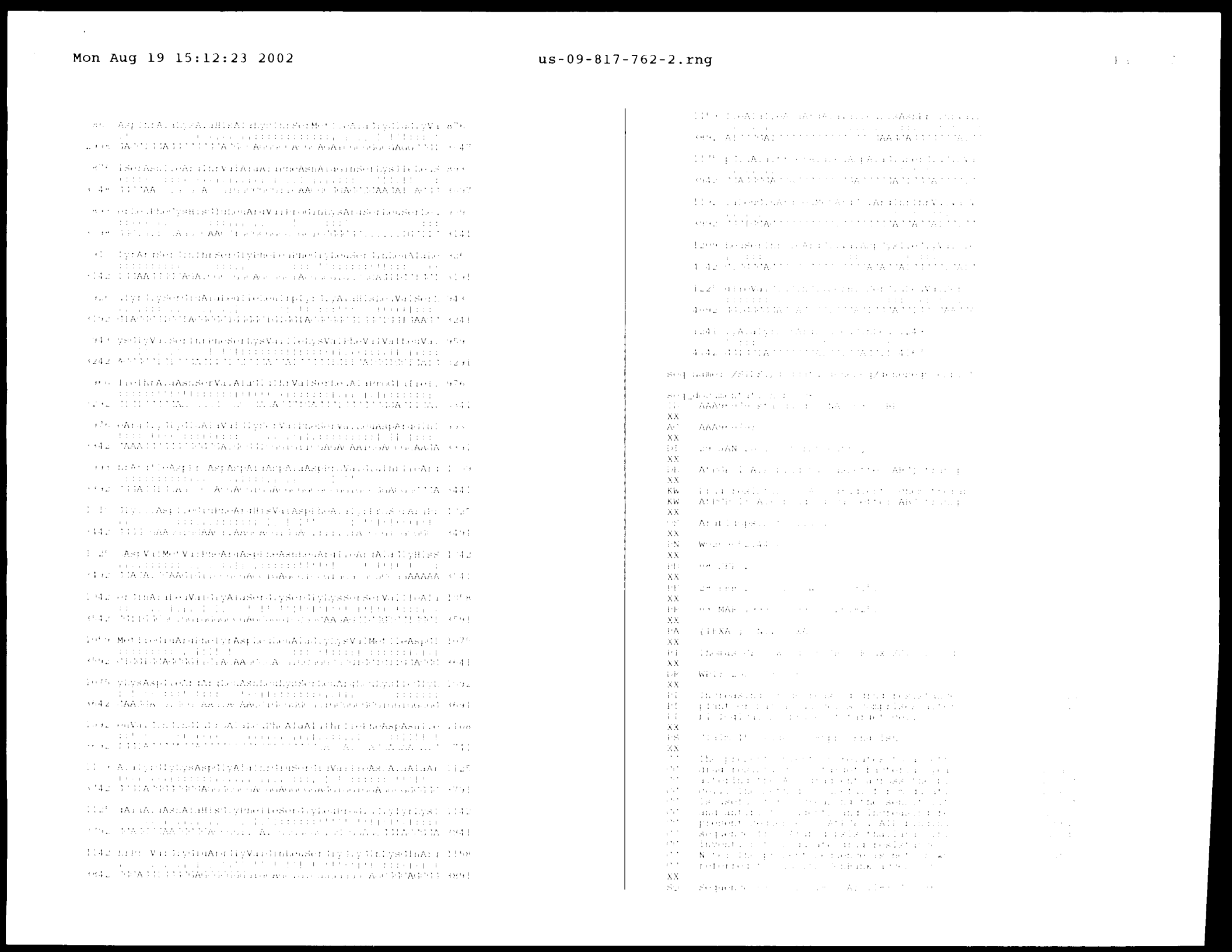


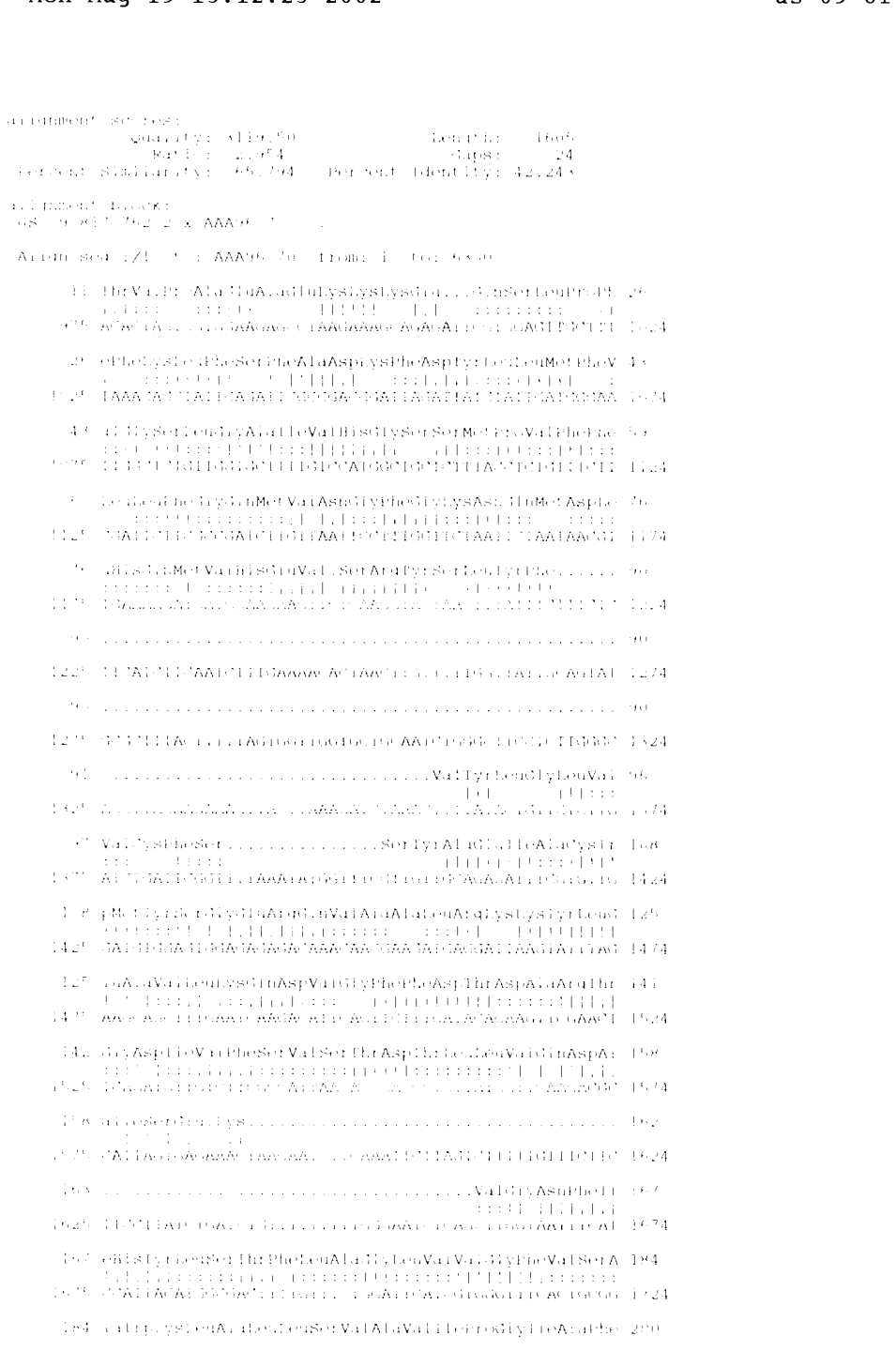
AC AAB47483;
XX
DT 09-MAY-2001 (first entry)
XX
DE Maize P-glycoprotein bz2 cDNA.
XX
KW bz2 gene; anti-fungal; Poaceae; Polycondylus transgenic plant;
KW disease resistance; P-g protein; fungicide.
XX
OS Zea mays.
XX
RN [1]
RP 1-4707
RX DOI; 10.1073/pnas.081065998.
RX PUBMED; 11353851.
RA Frame B.R., Drayton P.R., Bagnall S.V., Lewis D.J., Pegg S.E.,
RA Hussey C.E., Wang K.;
RT "Production of fertile transgenic maize plants by silicon carbide
RT whisker-mediated transformation";
RL Plant J. 31:205-220(2002).
XX
DR WPI; 2001-456789/49.
DR P-PSDB; AAB47483.
XX
FH Key Location/Qualifiers
FT source 1..4707
FT /organism="Zea mays"
FT /mol_type="genomic DNA"
FT /db_xref="taxon:4577"
FT /clone="BAC clone B73"
FT CDS 1..4707
FT /gene="bz2"
FT /product="P-glycoprotein"
FT /protein_id="AAB47483.1"
FT /translation="MAAQKNTLYLQMLGSLVAFFSSLQLGYNIGVMNAPQKVIEQDY
FT NISDVKRSVITSILEACFAAAPIFAAFLSDTYGRRPVILWSSAIFFIGAIIMGFSKL
FT AKSFEMMIIGRFIIGVYCGLTTGFVPMYVGEVSPTALRGALGTLHQLGIVVGILIAQ
FT VFGLDSIMGNKDLWPLLLSIIFIPALLQCIVLPFCPESPRFLLINRNEENRAKSVLKK
FT LRGTADVTHDMQEMIEEGRQMEKVPKATWGELFSTKPAVFQRVMMAMMIQSLQQLTG
FT INFFFYYGTTIFNAVGMKDSFETSIVLGVVNFASTFVSLYTVDRFGRRTCLLWGAVS
FT MMICGVIIASVGITRLWPHGQDQPSSKGAGNVMIVFACLFIFCFFFSWSPIAYVIVSE
FT LFLRGPRPAAFVVAGSVNWTSNFIVGMCFQYVEQLCGPYVFIIFTVLLVLFFIFIYL
FT RVPETKGRTFDEIASGFRQGGASQSDKTPEELFHPLGADSQV"
XX
SQ Sequence 4707 BP; 1411 A; 1005 C; 1054 G; 1237 T; 0 other;

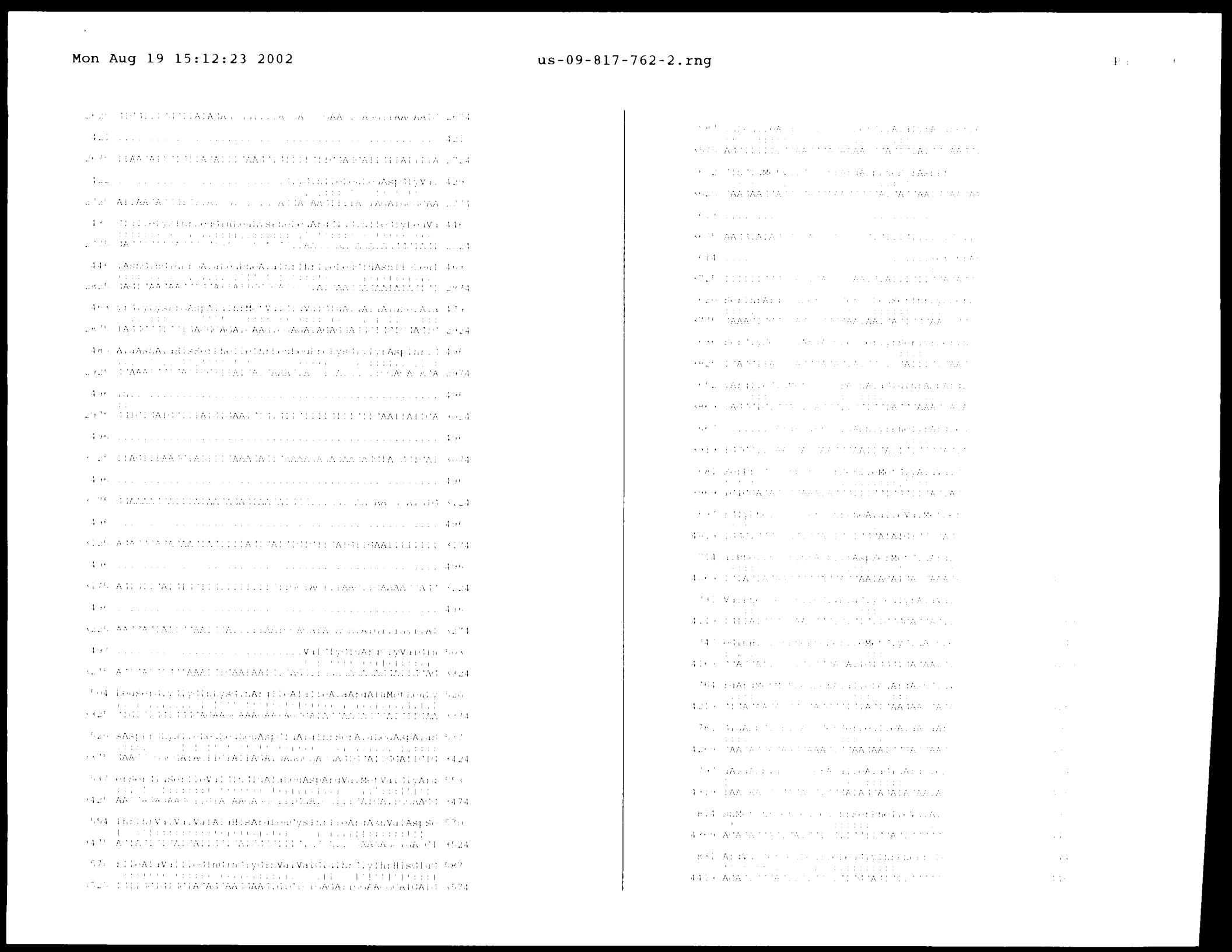
Alignment Scores:
Quality: 5842 Length: 1270

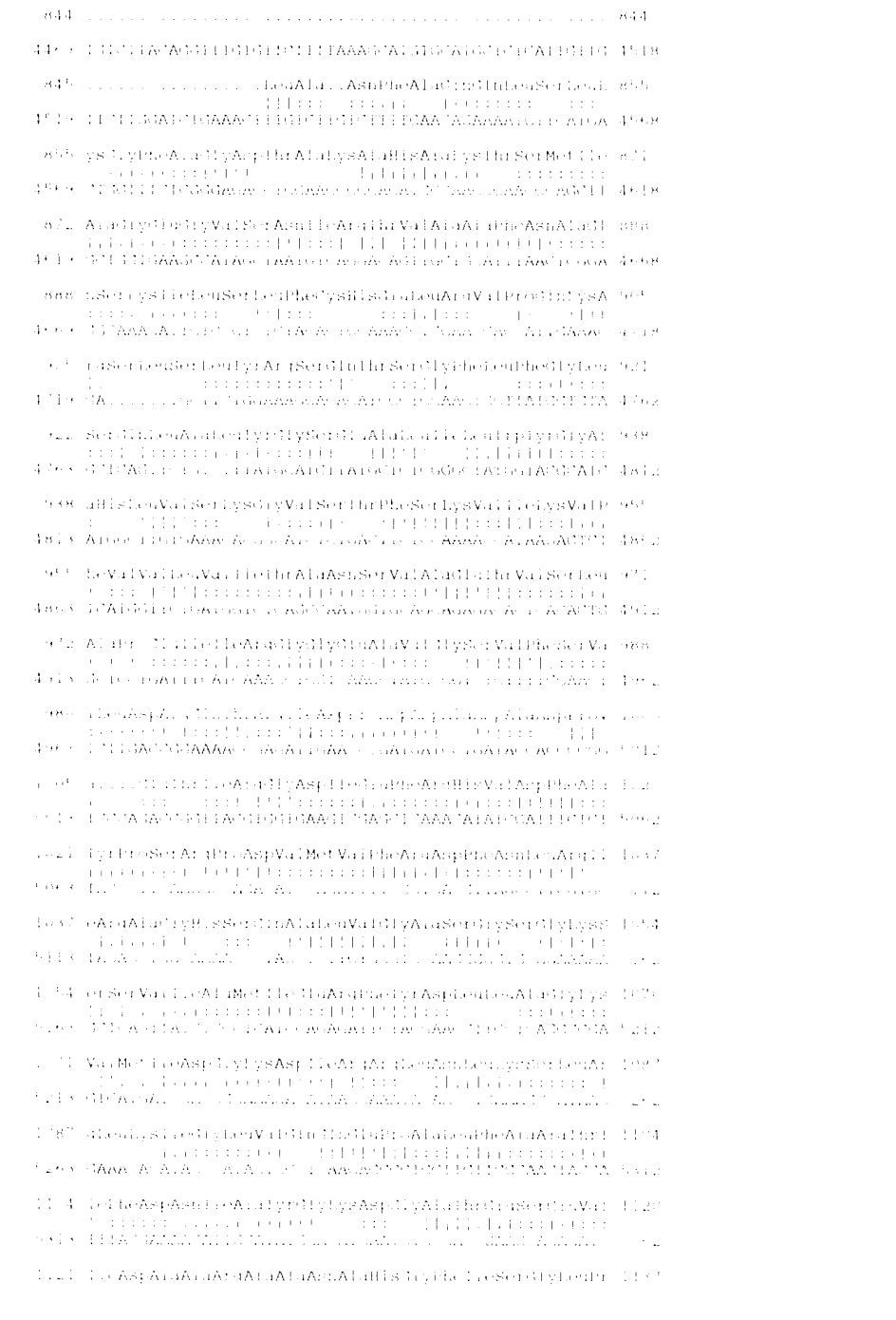


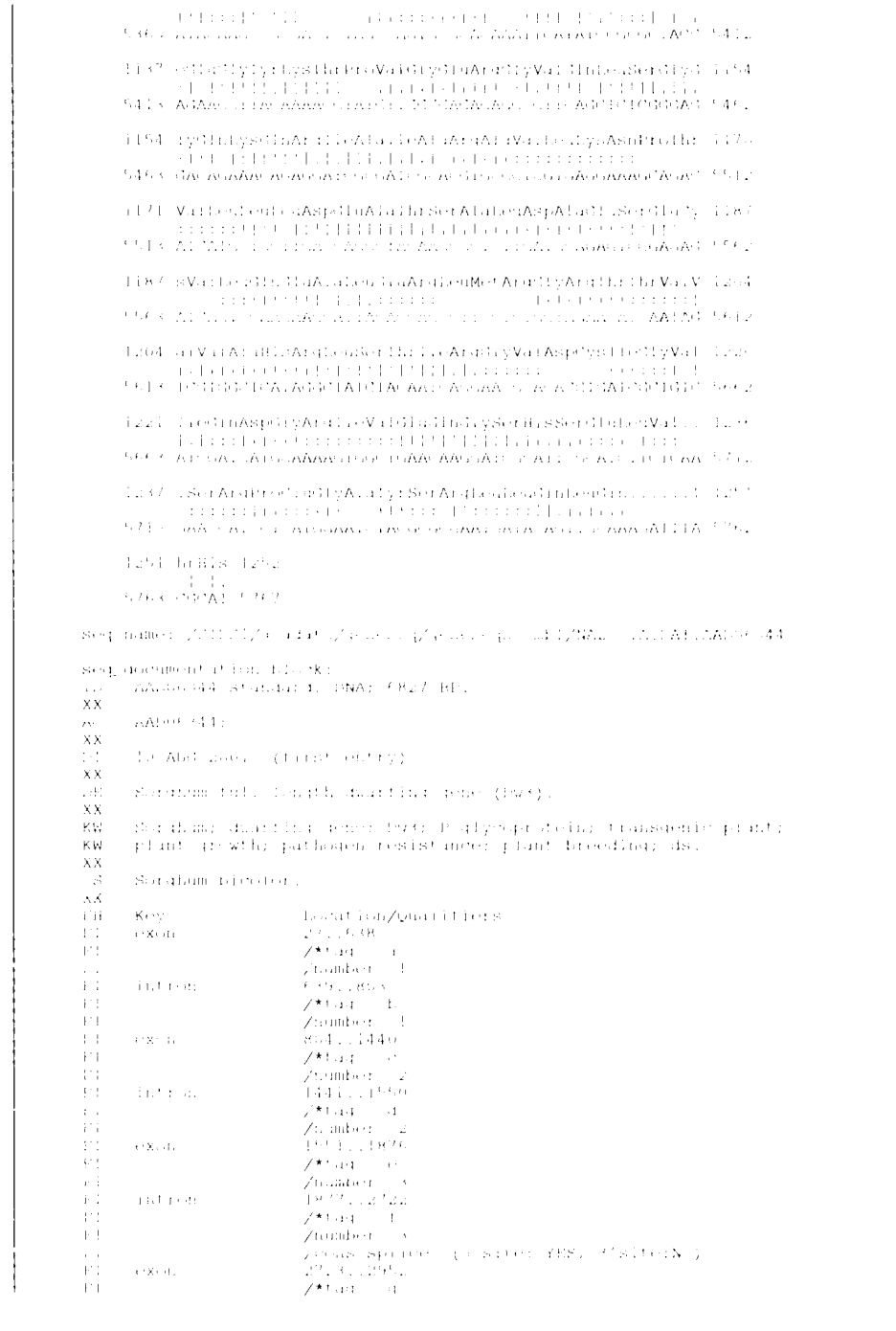


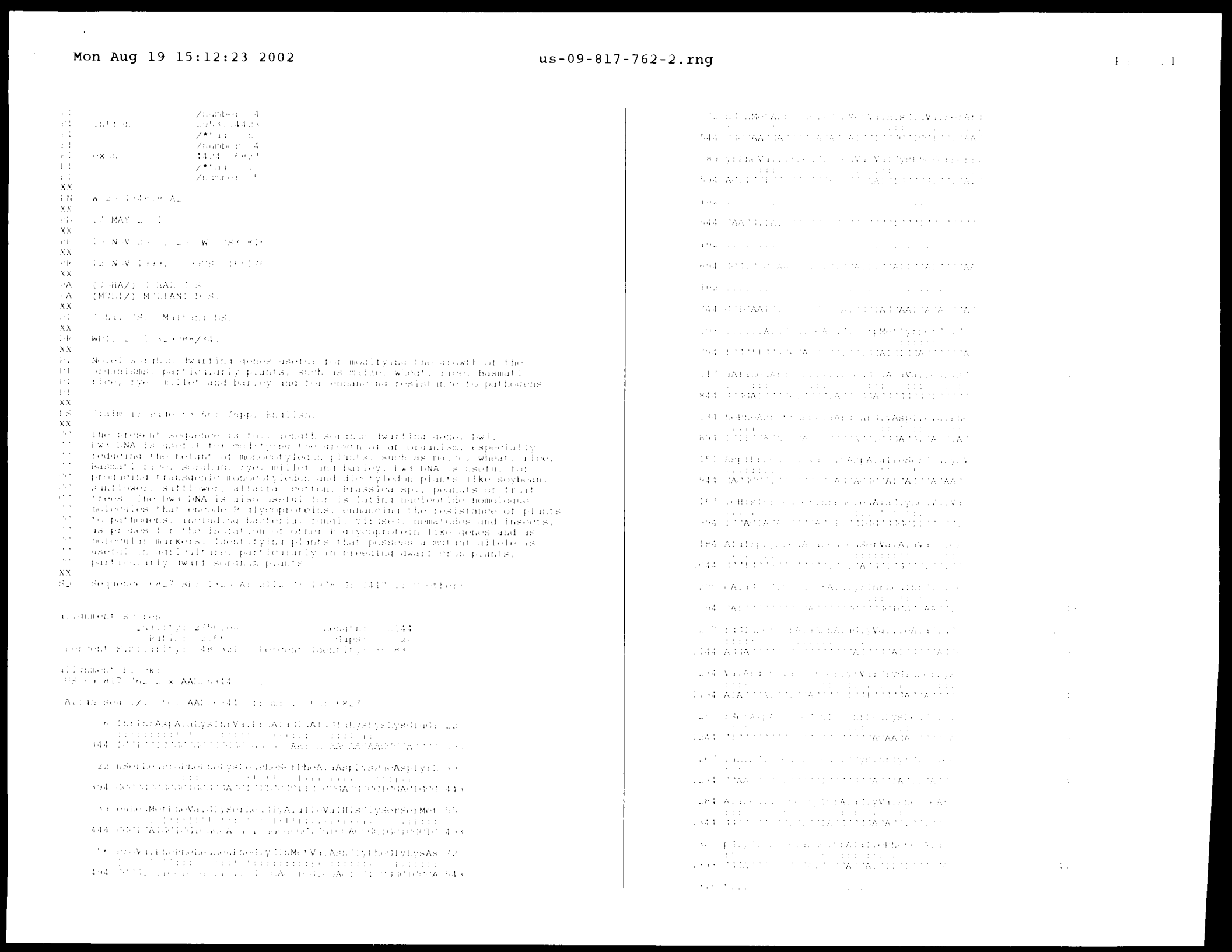


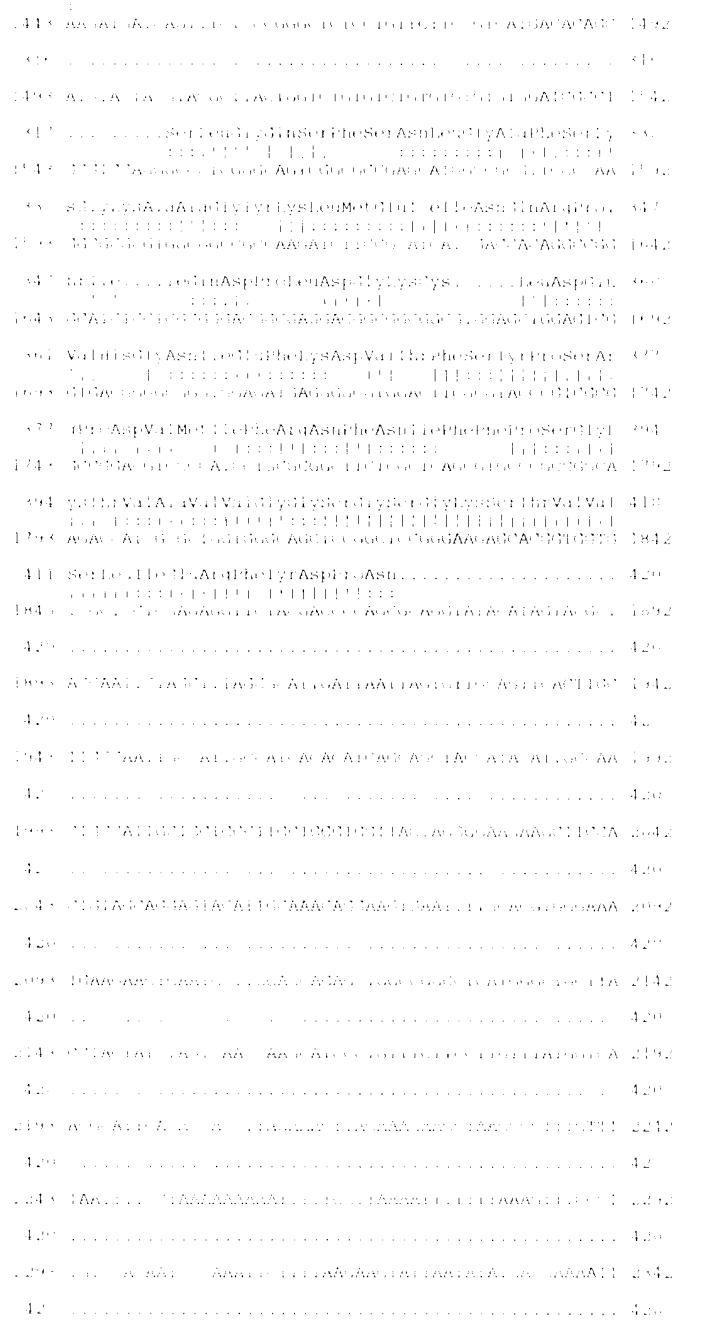


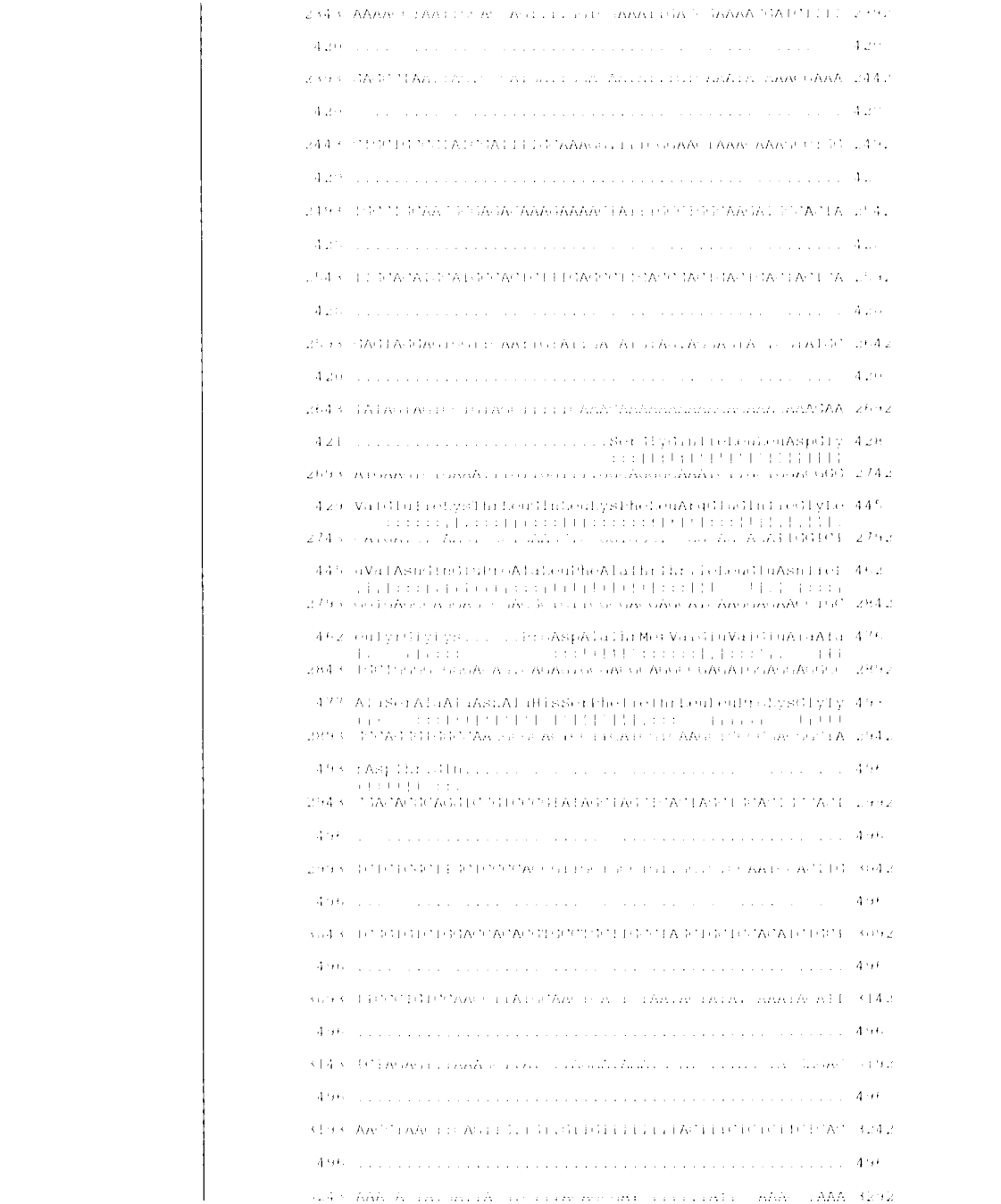


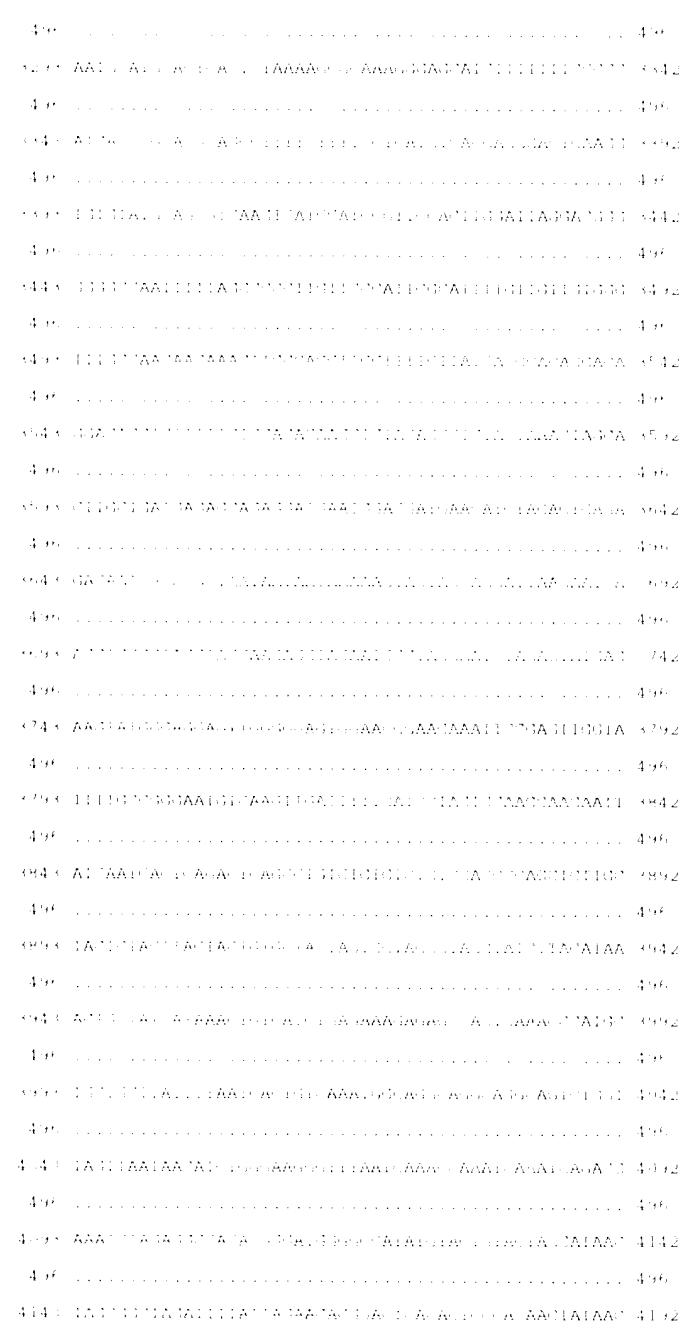


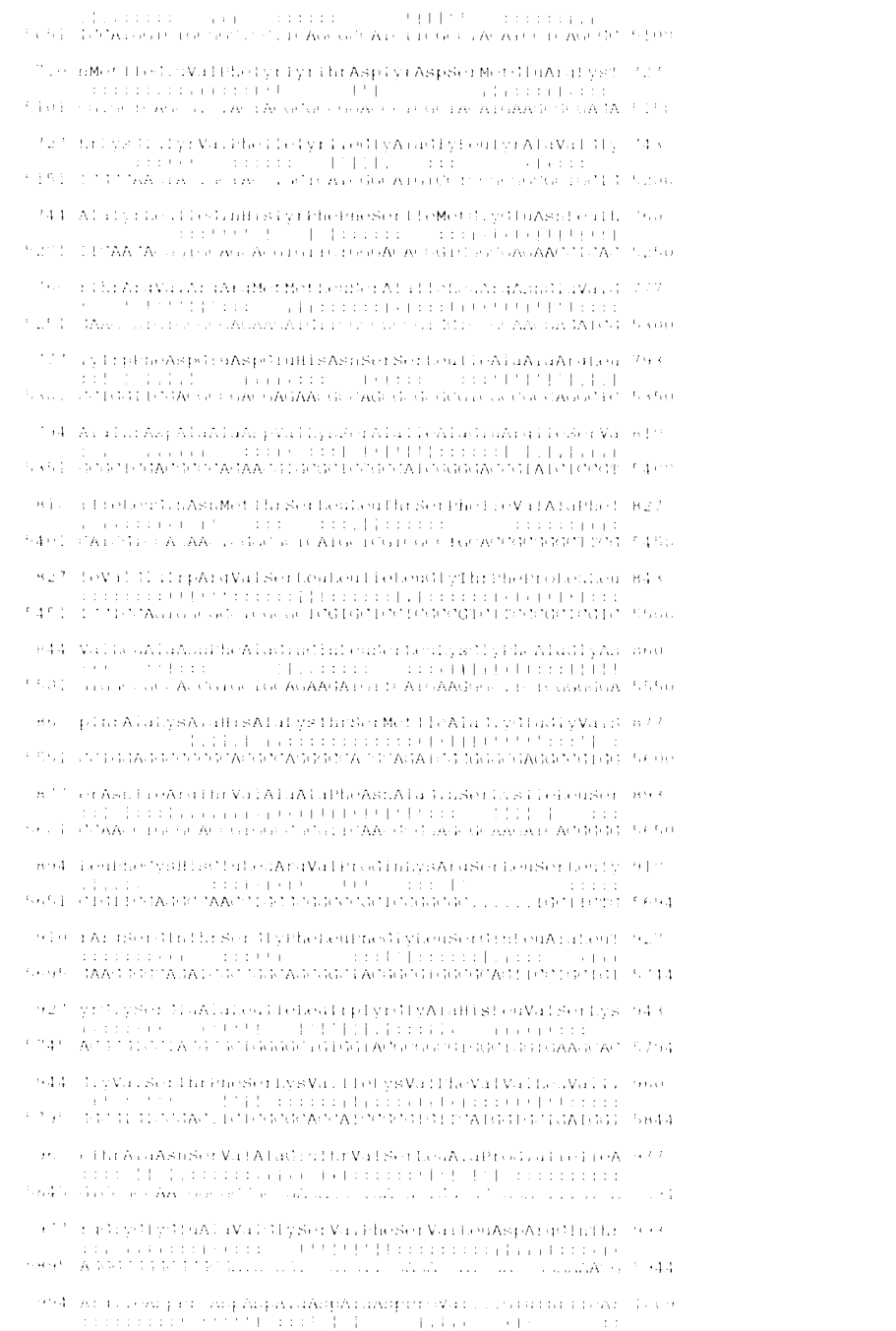


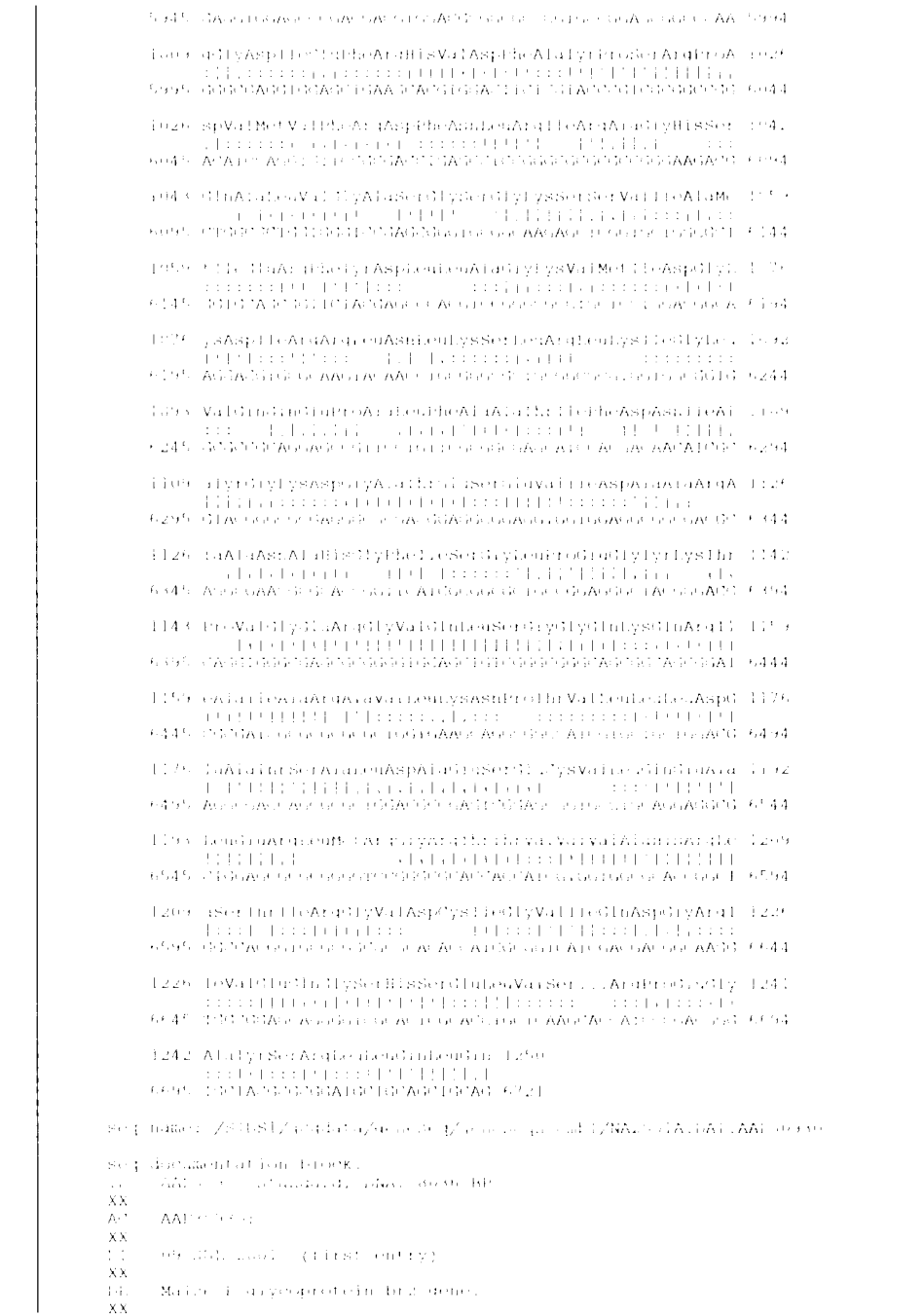












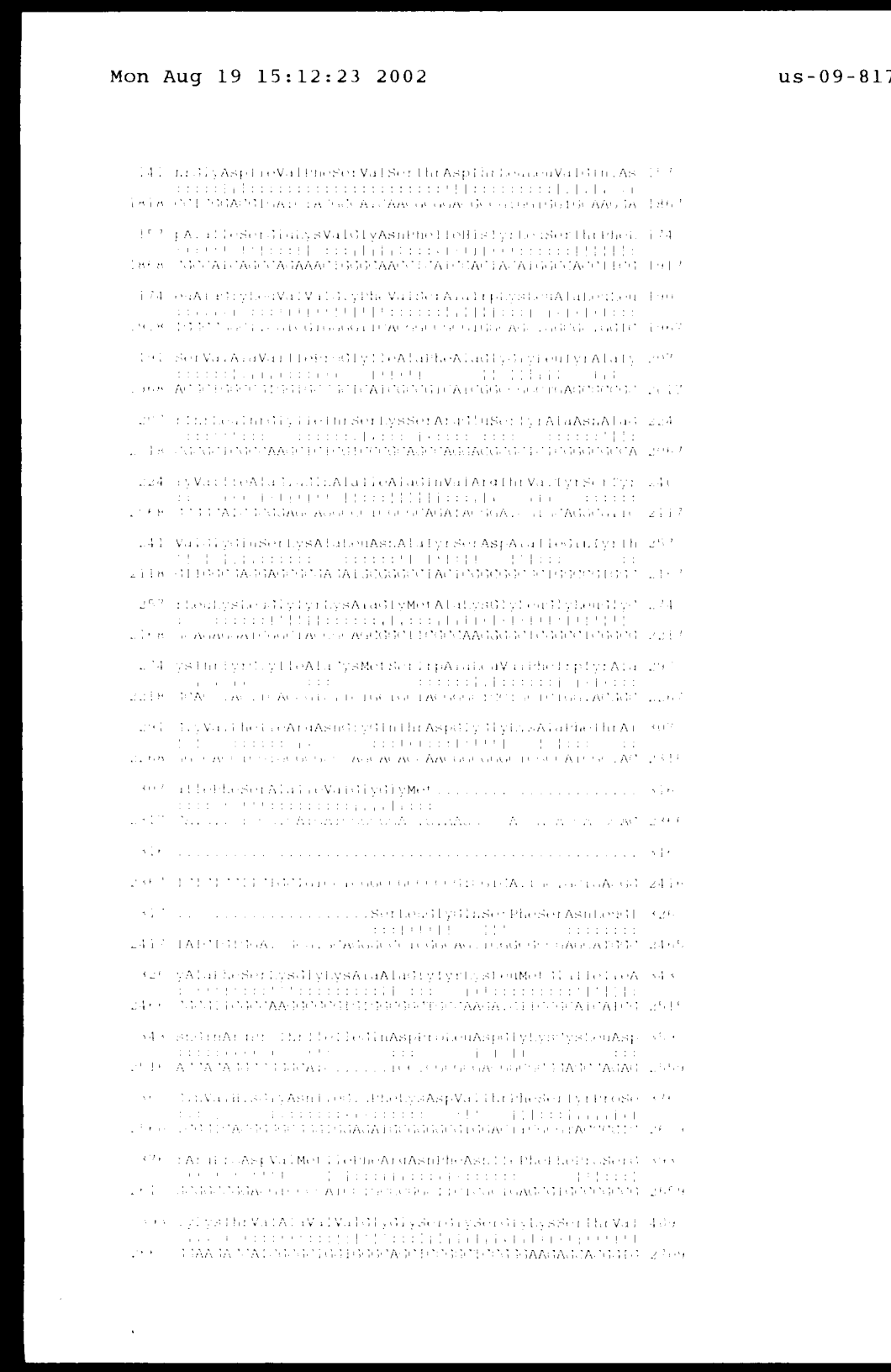
```
KW   Br2 gene; maize; [illegible] P-glycoprotein; transgenic plant;
KW   disease resistance; crop protection; dwarfing; ds.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             [illegible]
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            [illegible]..1562
FT                   /*tag= b
FT                   /number= "1"
FT   intron          [illegible]..1706
FT                   /*tag= c
FT                   /number= "1"
FT   exon            [illegible]..2434
FT                   /*tag= d
FT                   /number= "2"
FT   intron          [illegible]..24[illegible]
FT                   /*tag= e
FT                   /number= "2"
FT   exon            243[illegible]..274[illegible]
FT                   /*tag= f
FT                   /number= "3"
FT   intron          274[illegible]..[illegible]
FT                   /*tag= g
FT                   /number= "3"
FT   exon            [illegible]..[illegible]
FT                   /*tag= h
FT                   /number= "4"
FT   intron          [illegible]..5252
FT                   /*tag= i
FT                   /number= "4"
FT   exon            525[illegible]..8[illegible]
FT                   /*tag= j
FT                   /number= "5"
XX
PN   WO[illegible] A2.
XX
PD   17-MAY-[illegible].
XX
PF   [illegible]-NOV-[illegible]; [illegible]WO-US[illegible].
XX
PR   12-NOV-[illegible]; [illegible]US-[illegible].
XX
PA   ([illegible]HA/) [illegible]HA[illegible] [illegible] S.
PA   (MULT/) MULTANI D S.
PA   (BRIG/) BRIGGS S P.
XX
PI   [illegible] DS, Multani DS, Briggs SP;
XX
DR   WPI; 2001-[illegible]/34.
DR   P-PSDB; AAB2[illegible].
XX
PT   Novel isolated P-glycoprotein polynucleotide, preferably br2 gene of
PT   maize, useful for modifying the growth of organisms, particularly
PT   plants and for enhancing resistance of plants to pathogens
XX
PS   Claim [illegible]; Page [illegible]; [illegible]pp; English.
XX
CC   The present sequence is of the maize br2 gene, which encodes a
CC   1[illegible] amino acid P-glycoprotein (see AAB20497). The gene is located
CC   on maize chromosome 1L. It was cloned by Mutator transposon
CC   tagging. Sequencing revealed extensive sequence and structural
CC   similarity with the multidrug resistance (MDR)-like gene encoded
CC   P-glycoproteins. BR2 may be involved in plant defence responses
CC   against pathogens. br2 cDNA (see AAF[illegible]) was also isolated.
CC   The invention provides compositions and methods for expressing
CC   genes encoding P-glycoproteins, such as maize br2, in plants. The
CC   br2 nucleic acids are useful for transforming plants such as corn,
CC   rice, soybean, olive, coconut, sugarcane, oats, barley, vegetables,
CC   ornamentals, and conifers for tissue preferred or constitutive
CC   expression of P-glycoprotein, for isolating homologous nucleotide
CC   [illegible]
CC   resistance [illegible]
CC   the growth [illegible]
CC   provided [illegible]
CC   alleles [illegible]
CC   an organism [illegible]
CC   the breeding [illegible]
CC   transformed [illegible]
CC   provided. [illegible]
CC   expression [illegible]
CC   probes [illegible]
CC   molecular [illegible]
CC   height [illegible]
CC   alternatives [illegible]
CC   growth [illegible]
CC   [illegible]
XX
SQ   Sequence [illegible]

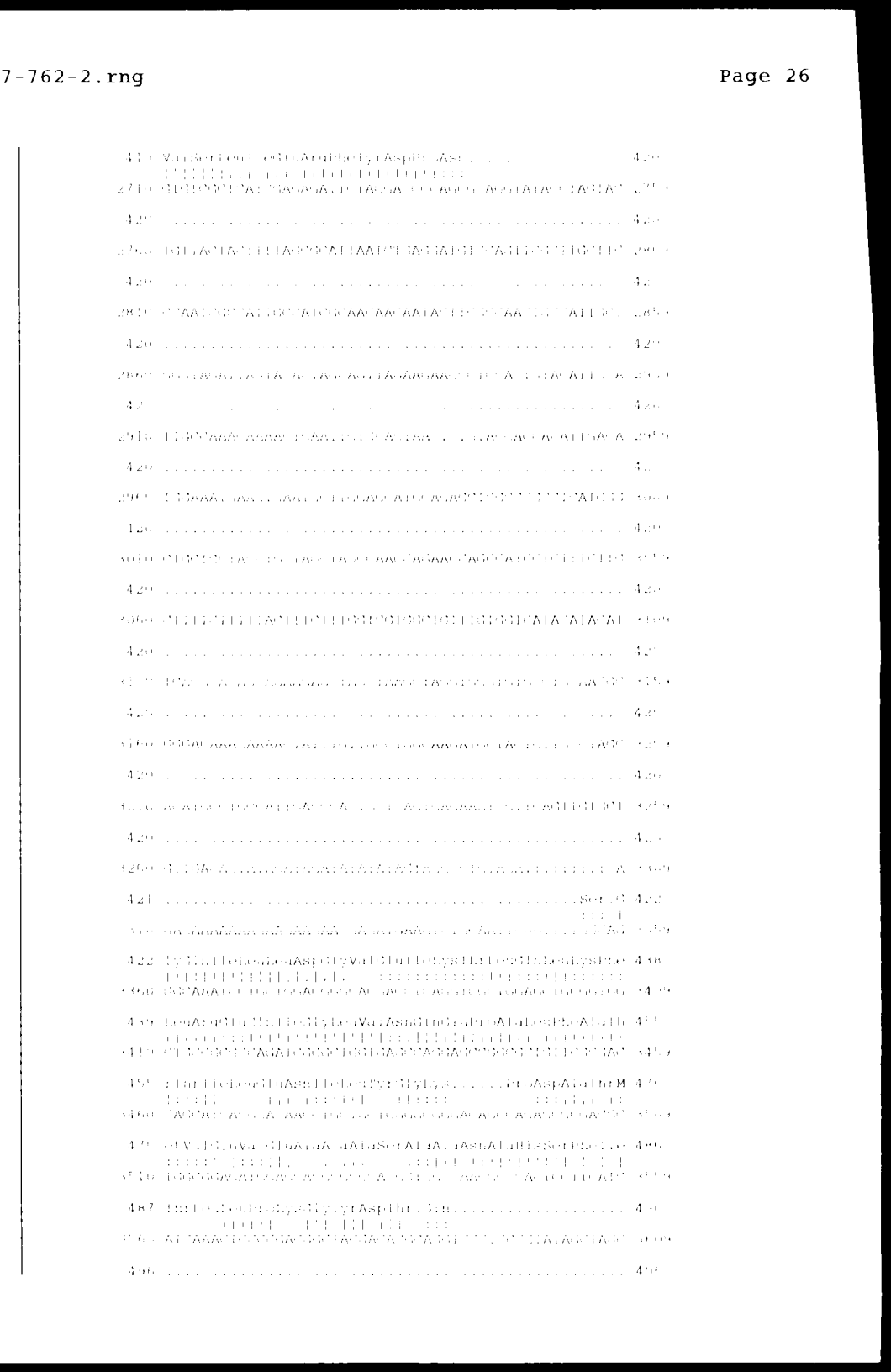
  Alignment Scores:
  [illegible]
  Percent Similarity: [illegible]

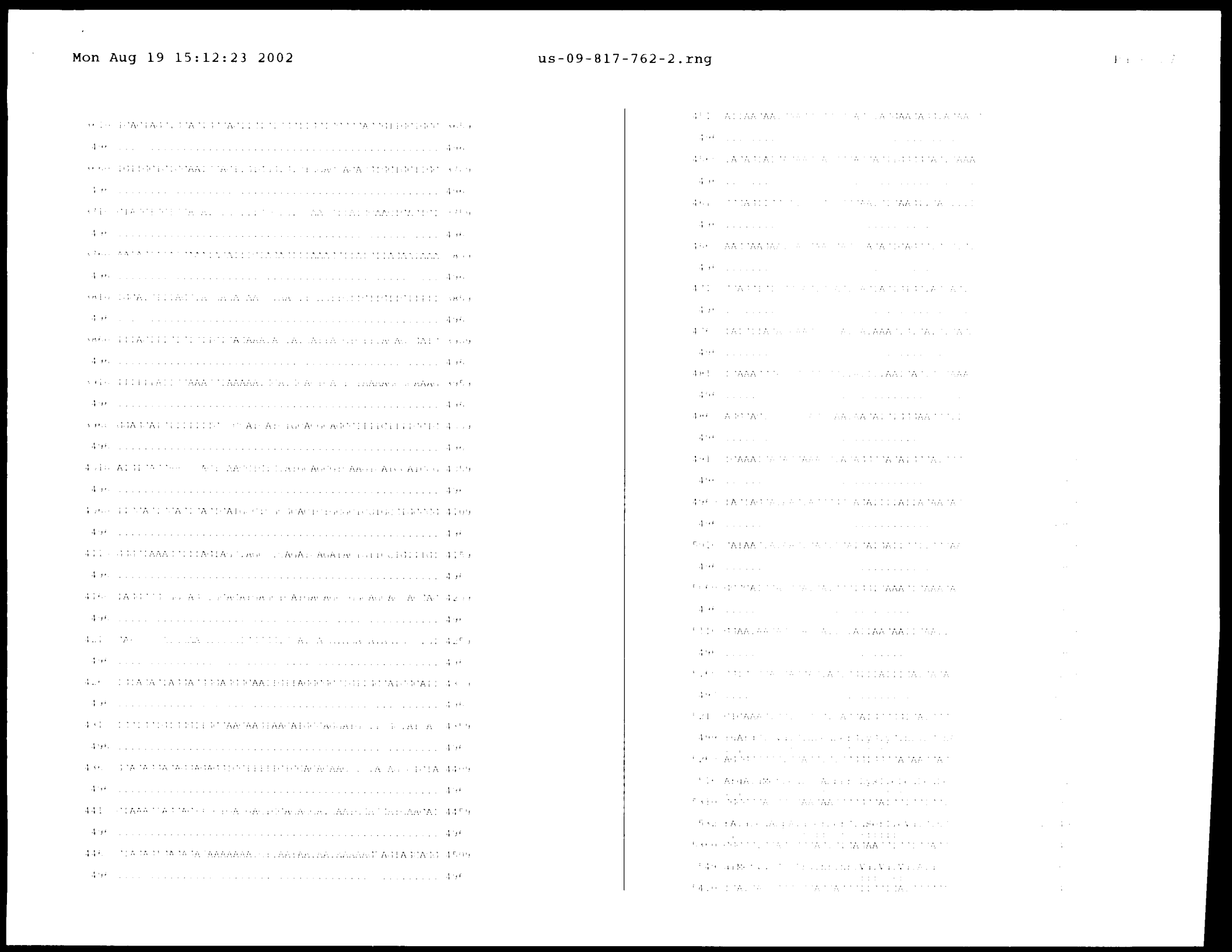
  Alignment to Db:
  US-09-817-762-2 [illegible]

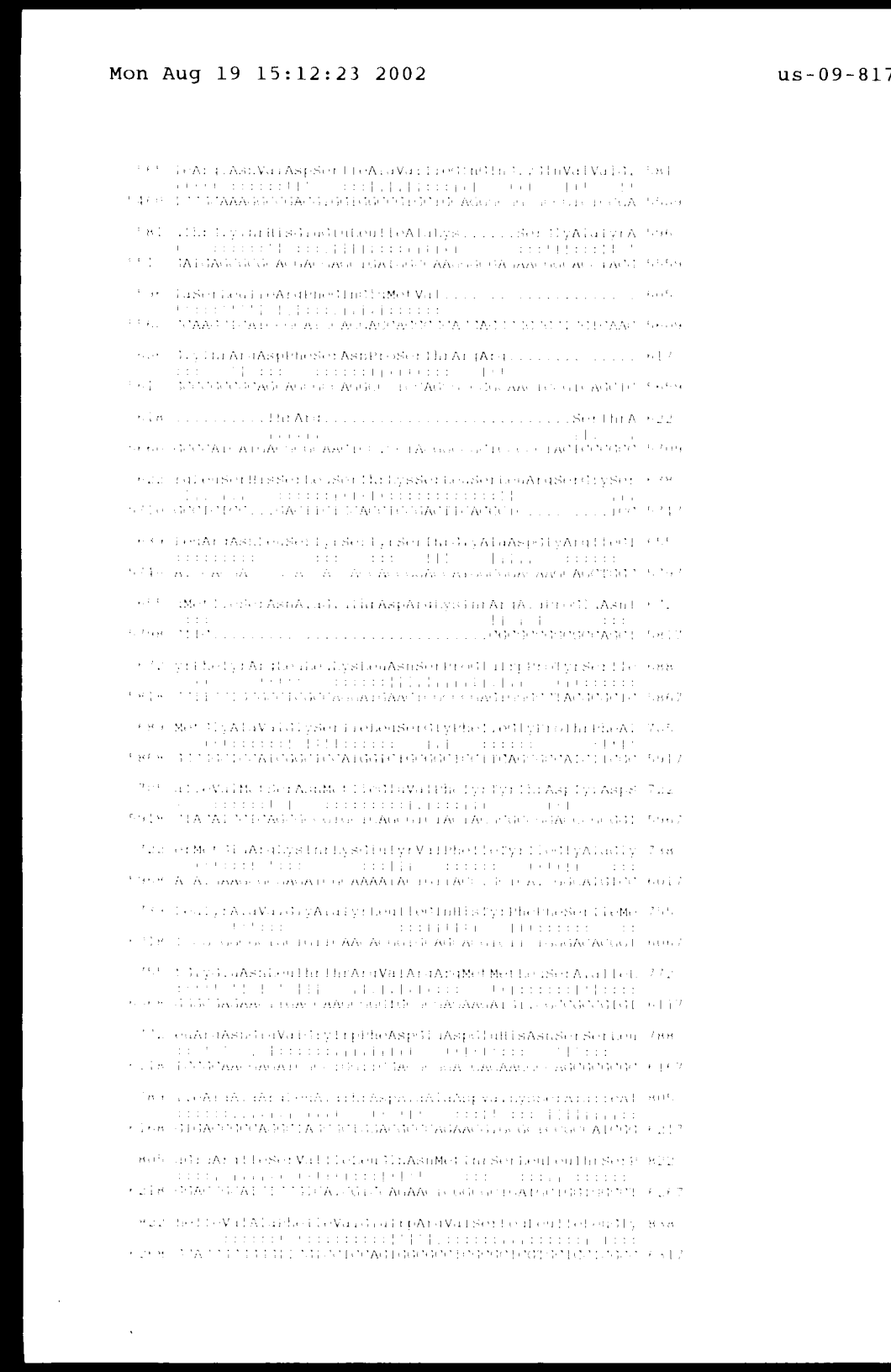
  Align seq 1/1 to: [illegible]
```

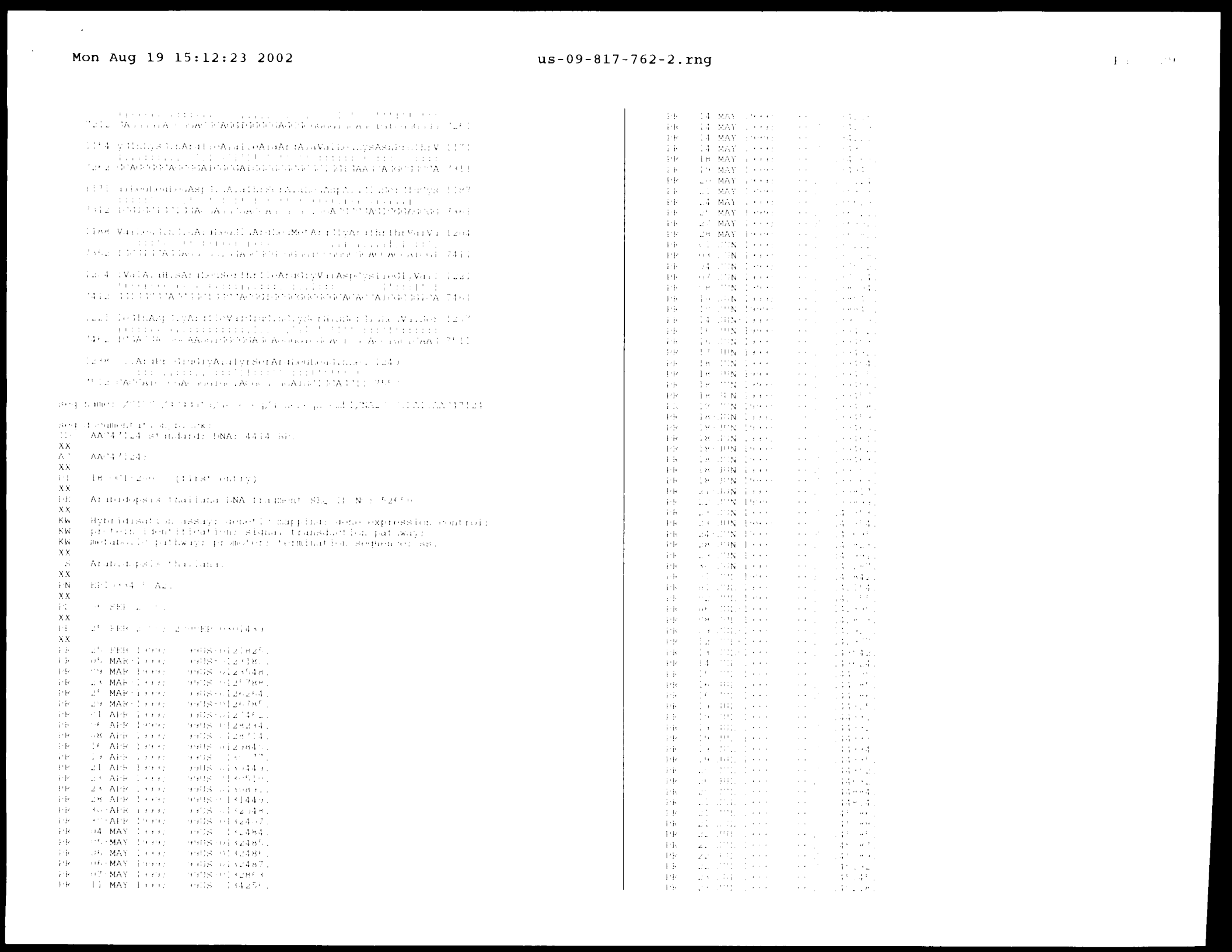
[illegible]

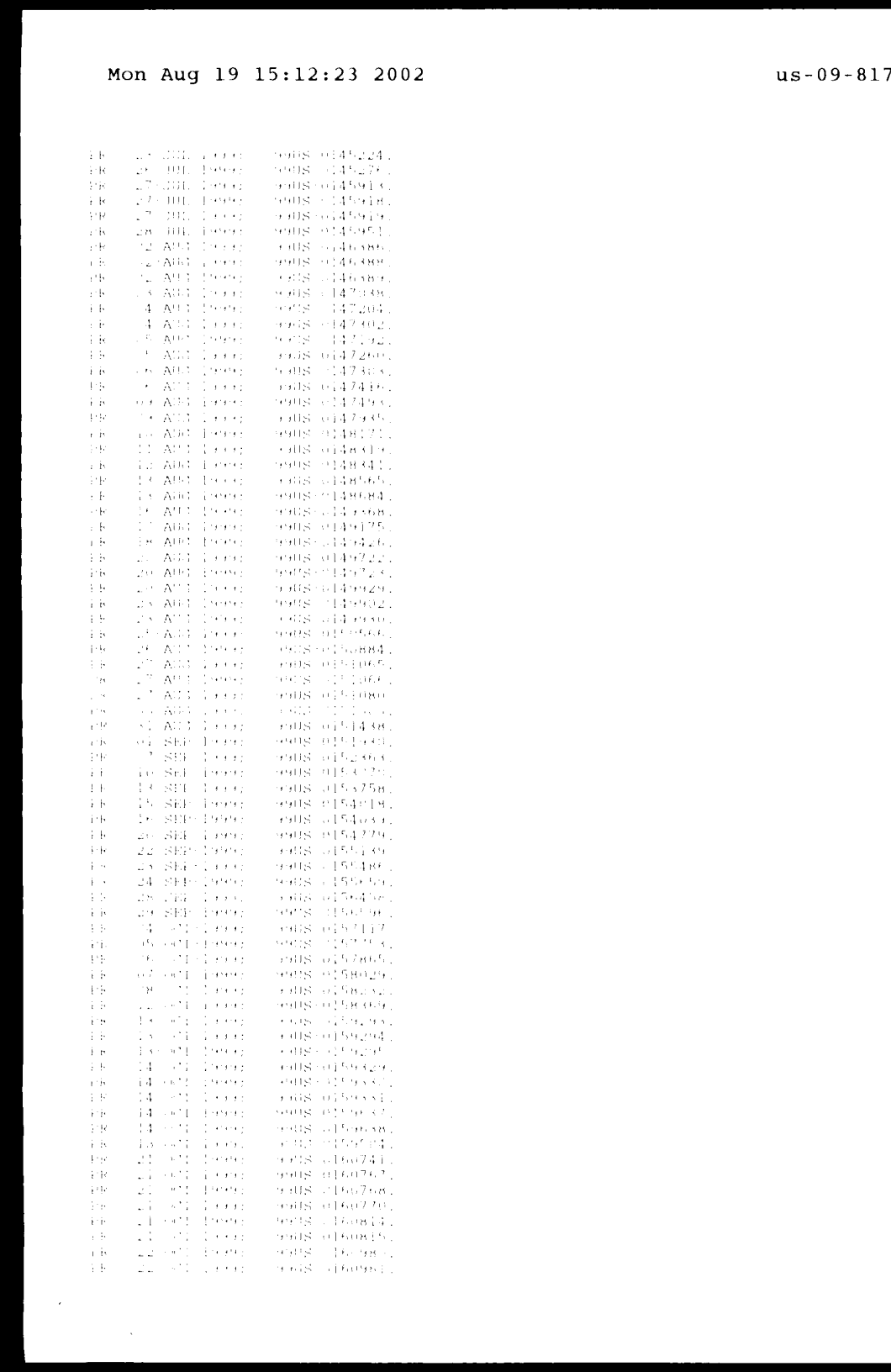


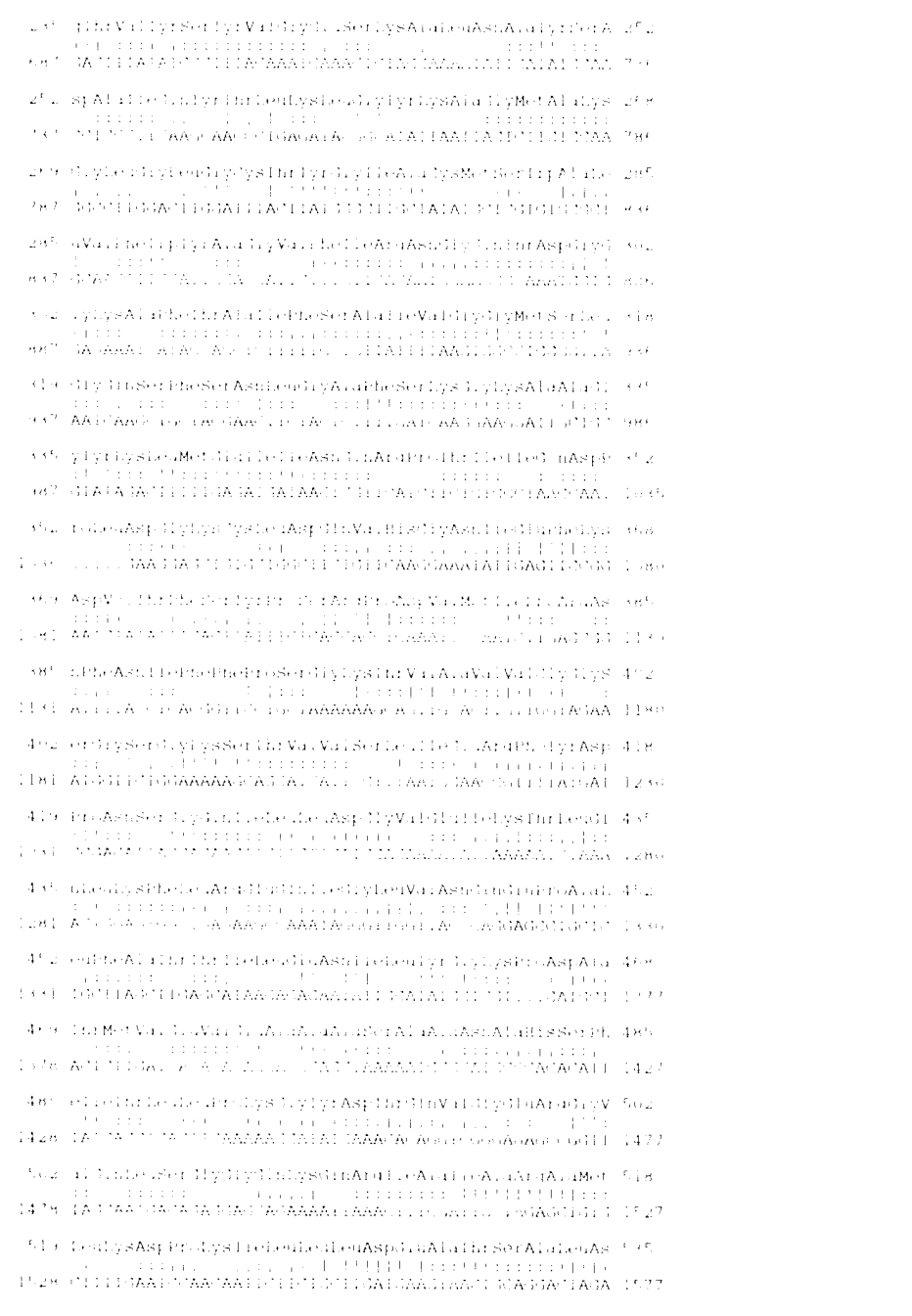


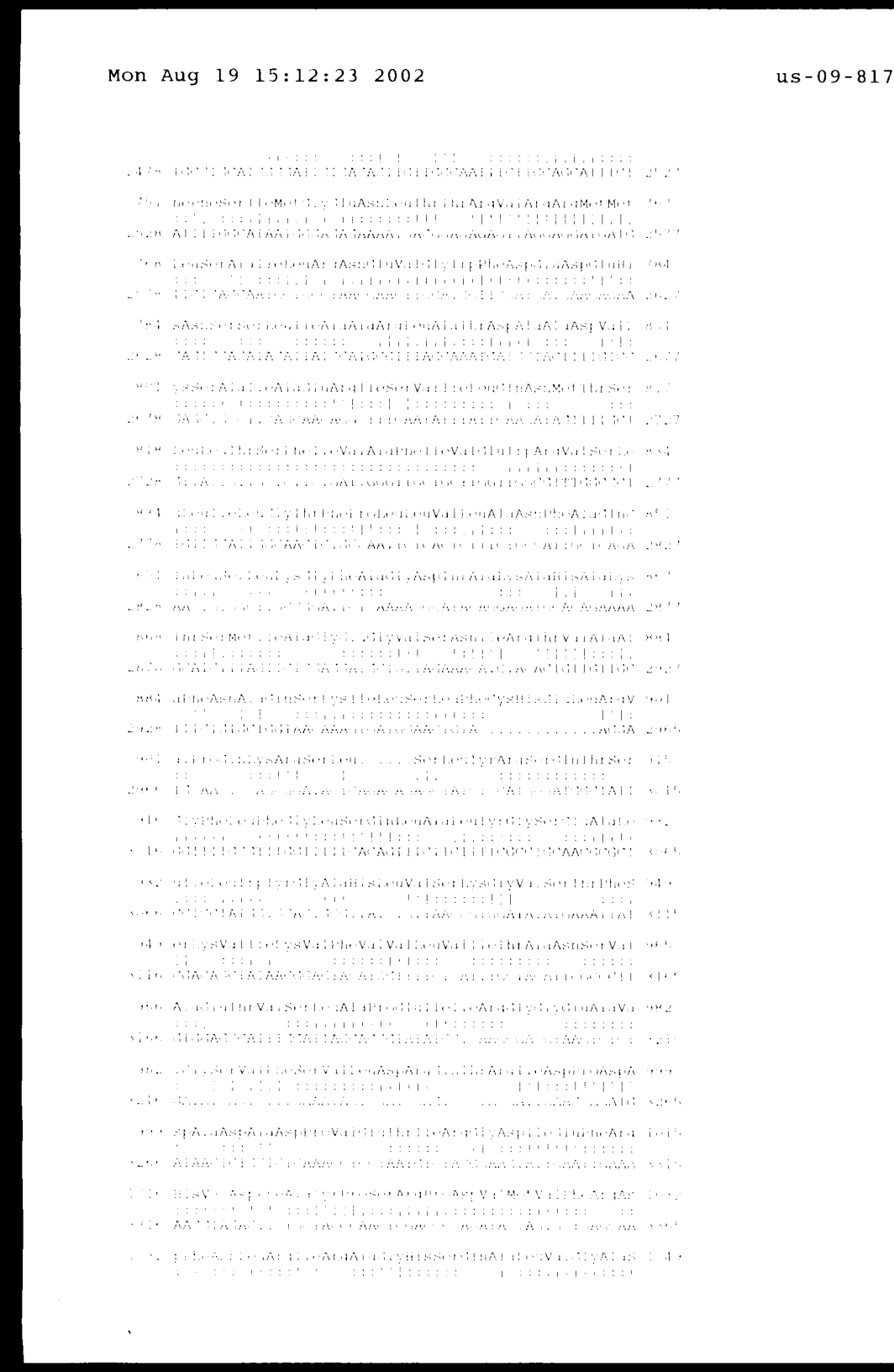




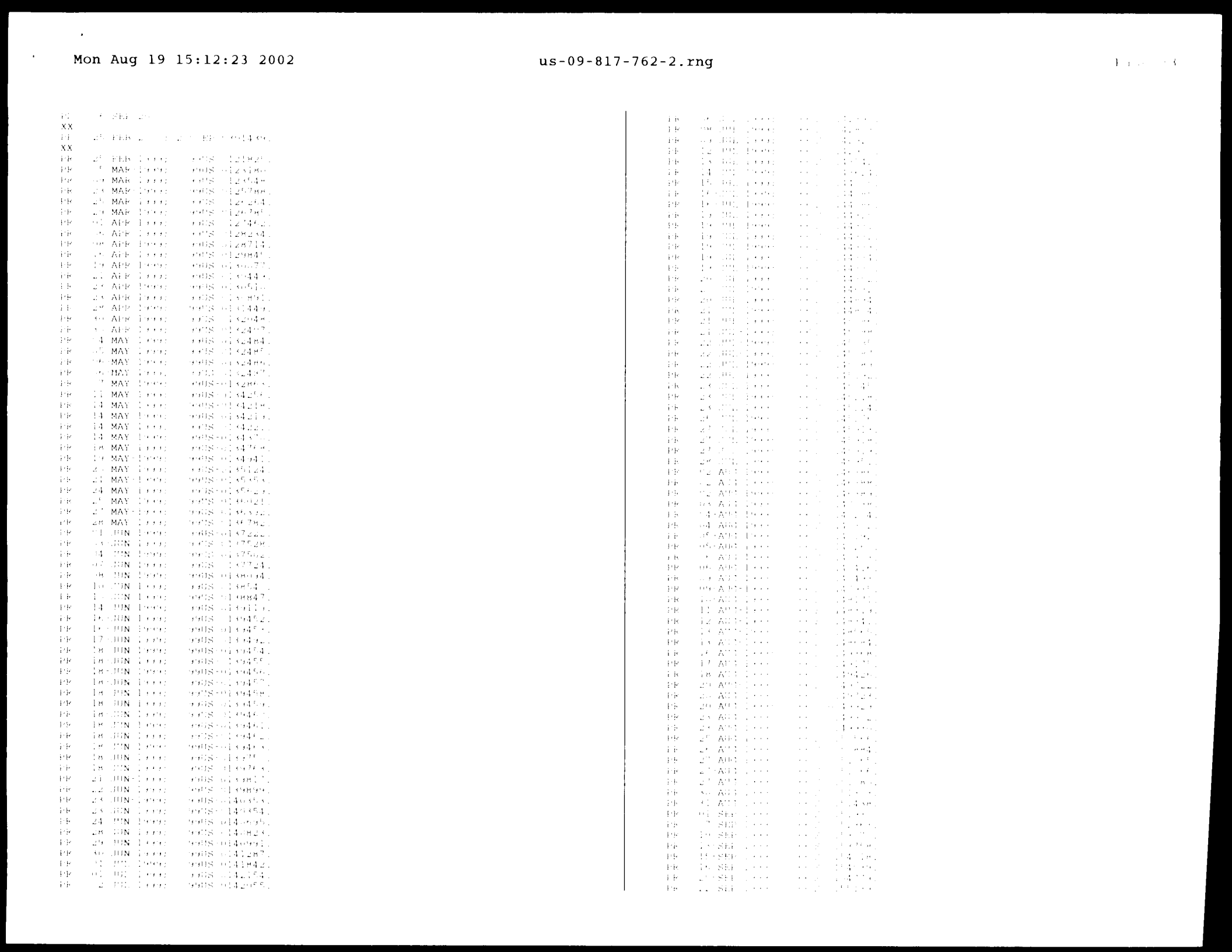


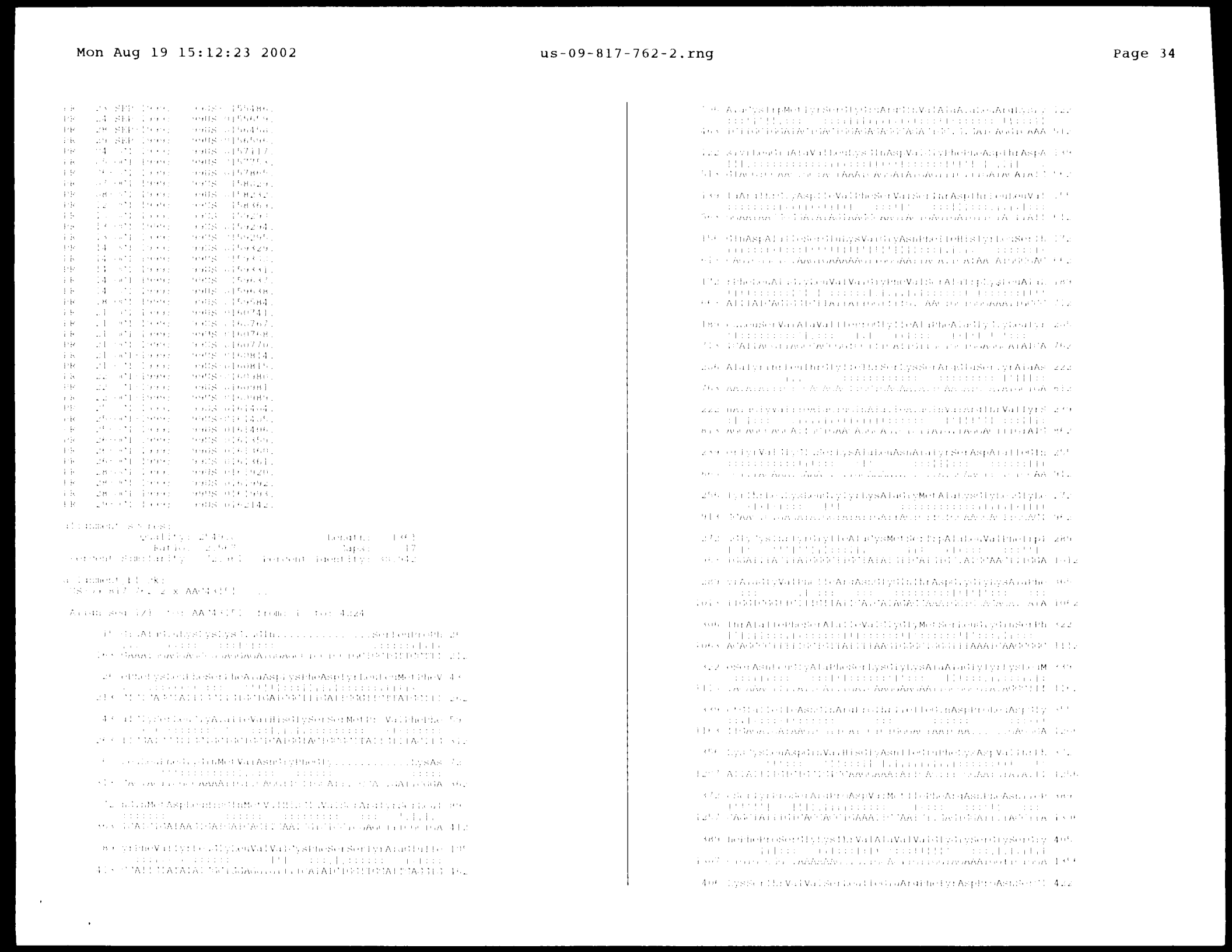


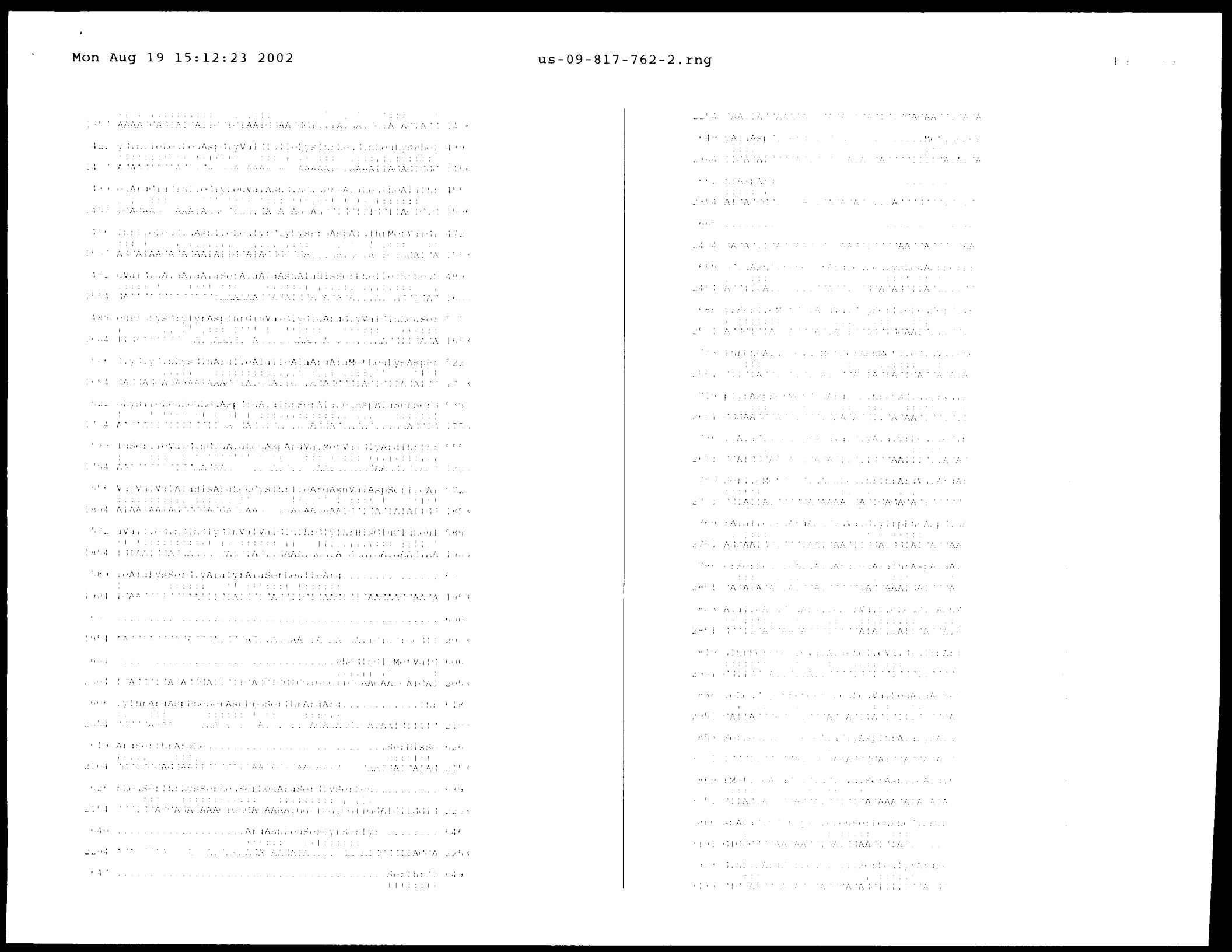


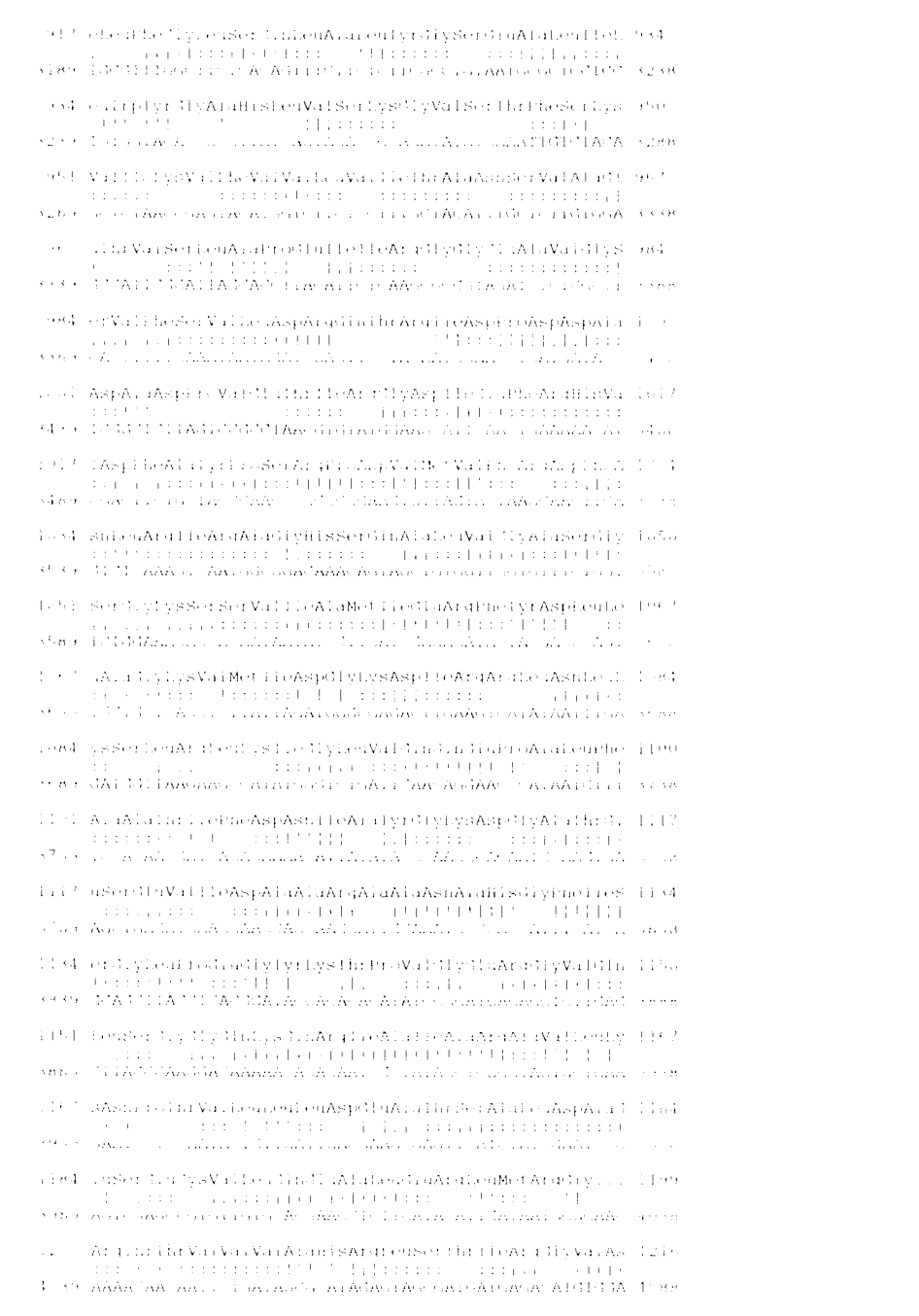


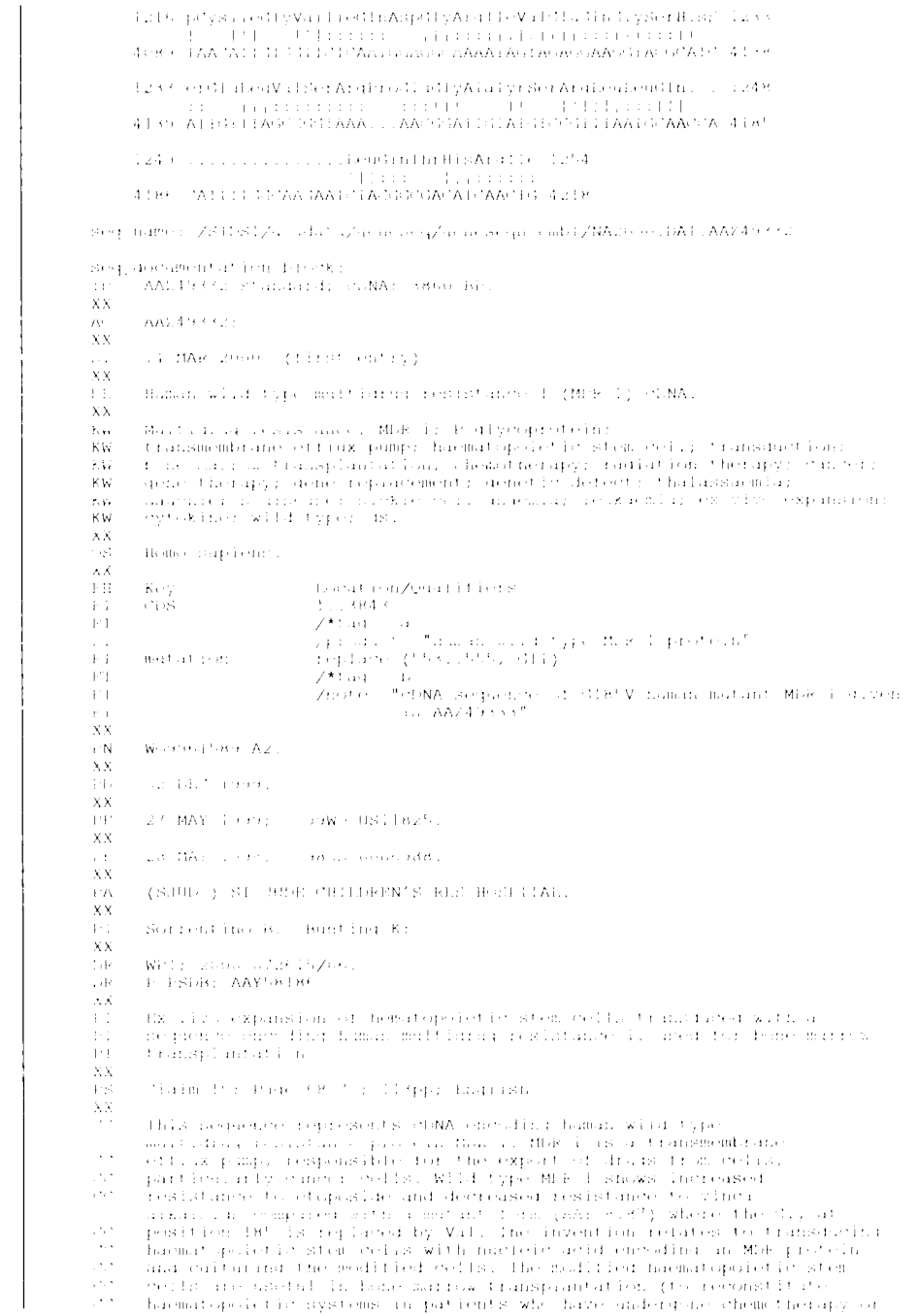
Seq_database/division_code:
ID XXX; XXX; 1224 BP.
XX
AC XXX;
XX
DT 17-OCT-2000 (Rel. 65, Created)
XX
DE Arabidopsis thaliana DNA fragment 270, 10 N : 98214.
XX
KW Hypothetical protein; assay; identification; gene expression; cDNA;
KW predicted; identification; signal transduction pathway;
KW transcription; promoter; functional; sequence.
XX
OS Arabidopsis thaliana
XX
OC EUK; xxx; Xxx.
XX

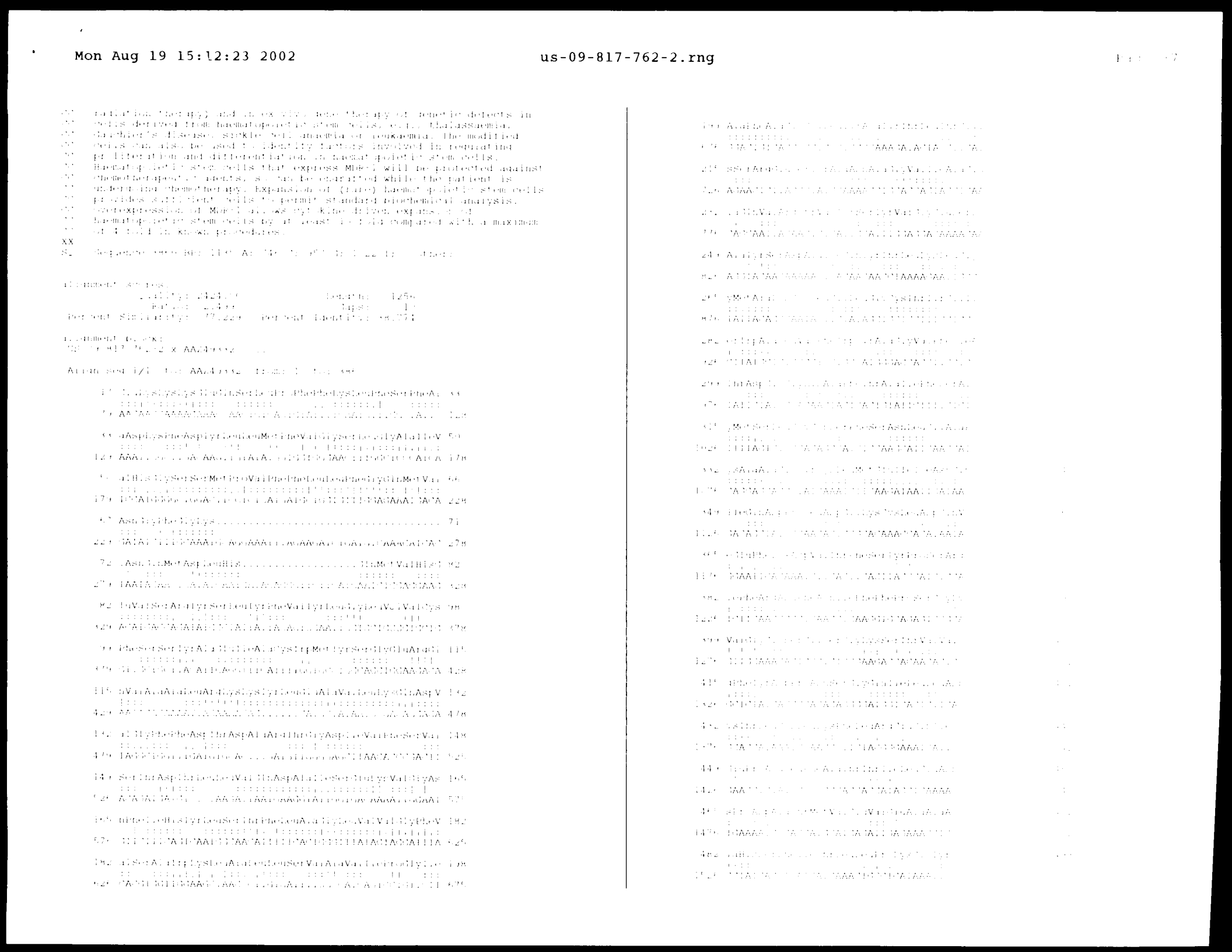


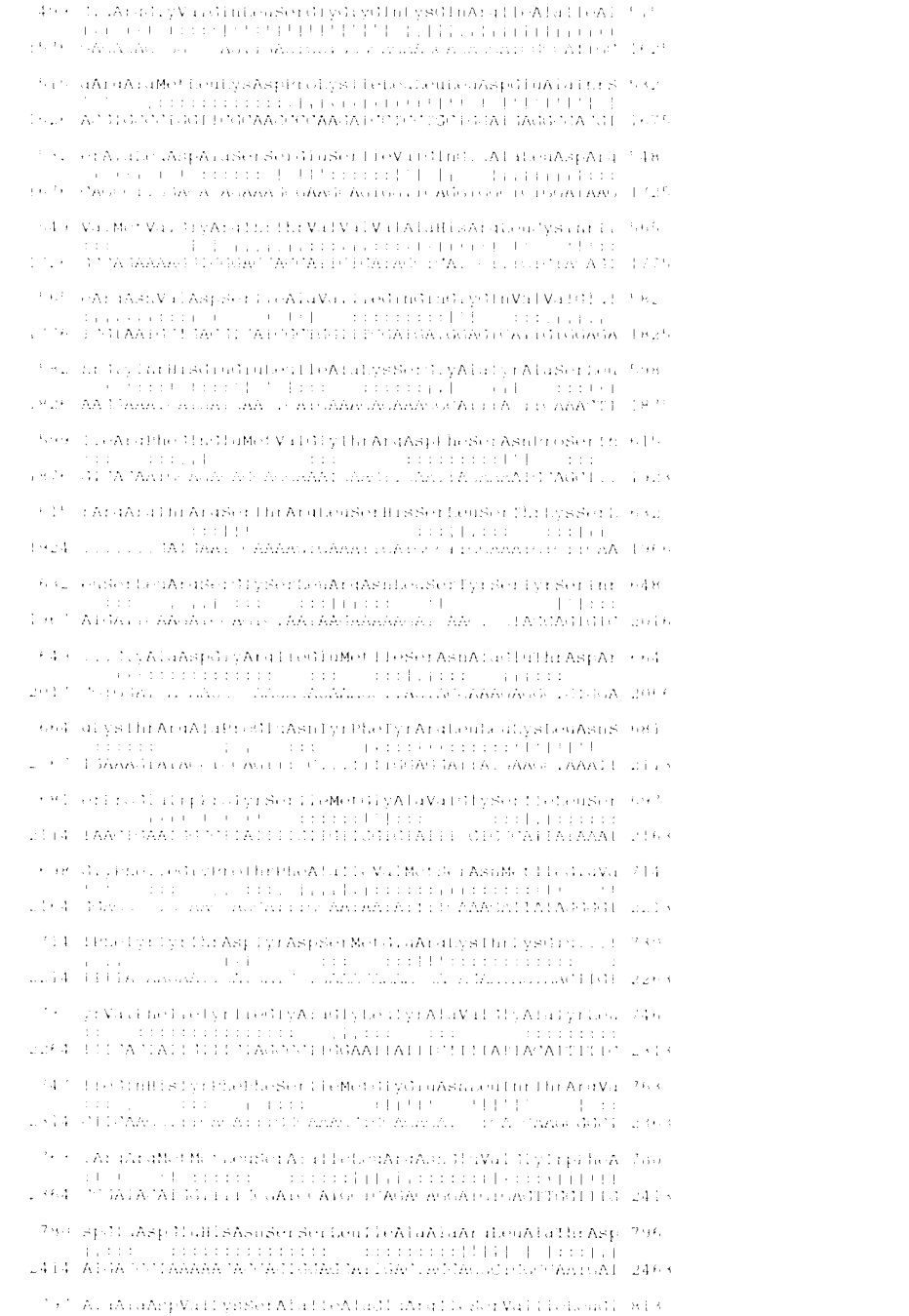


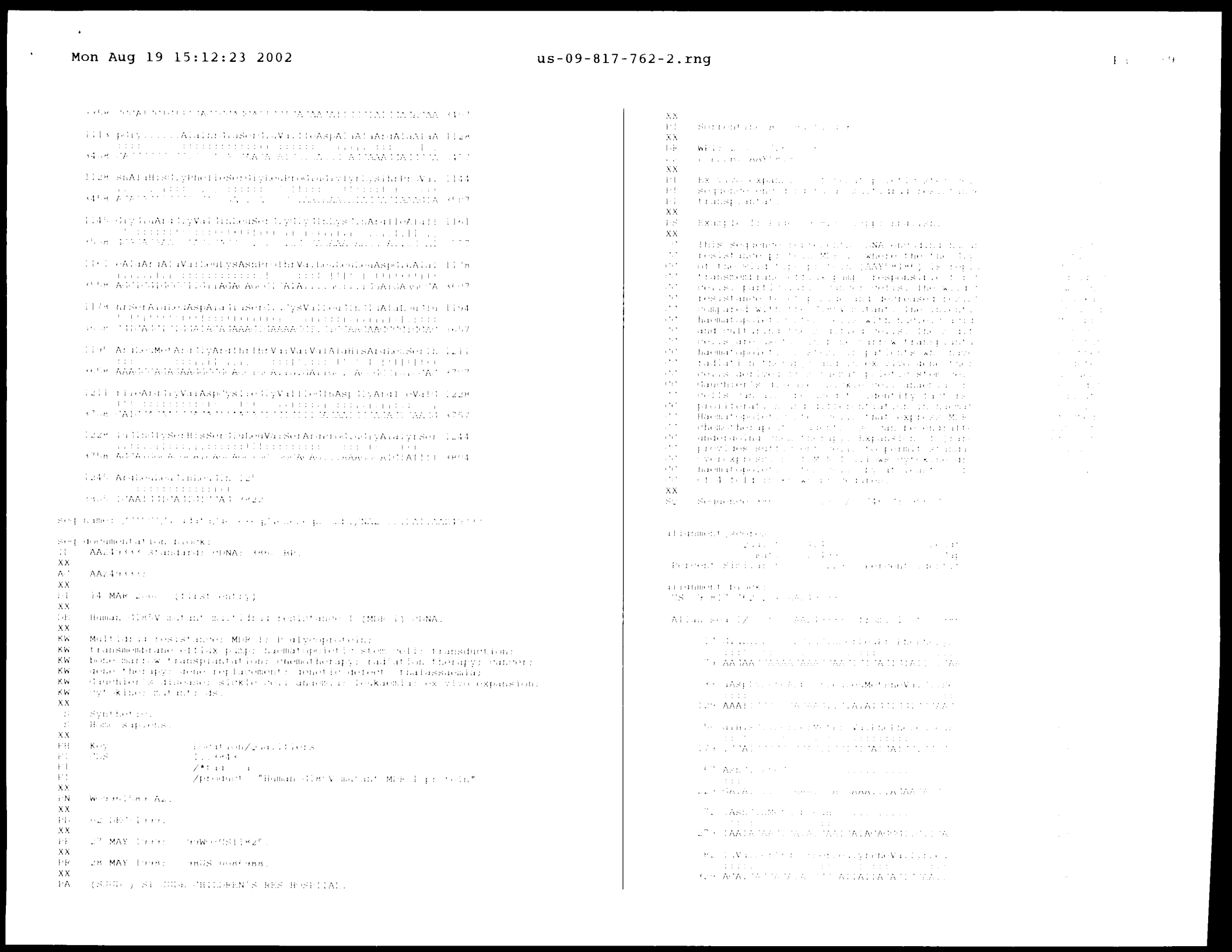


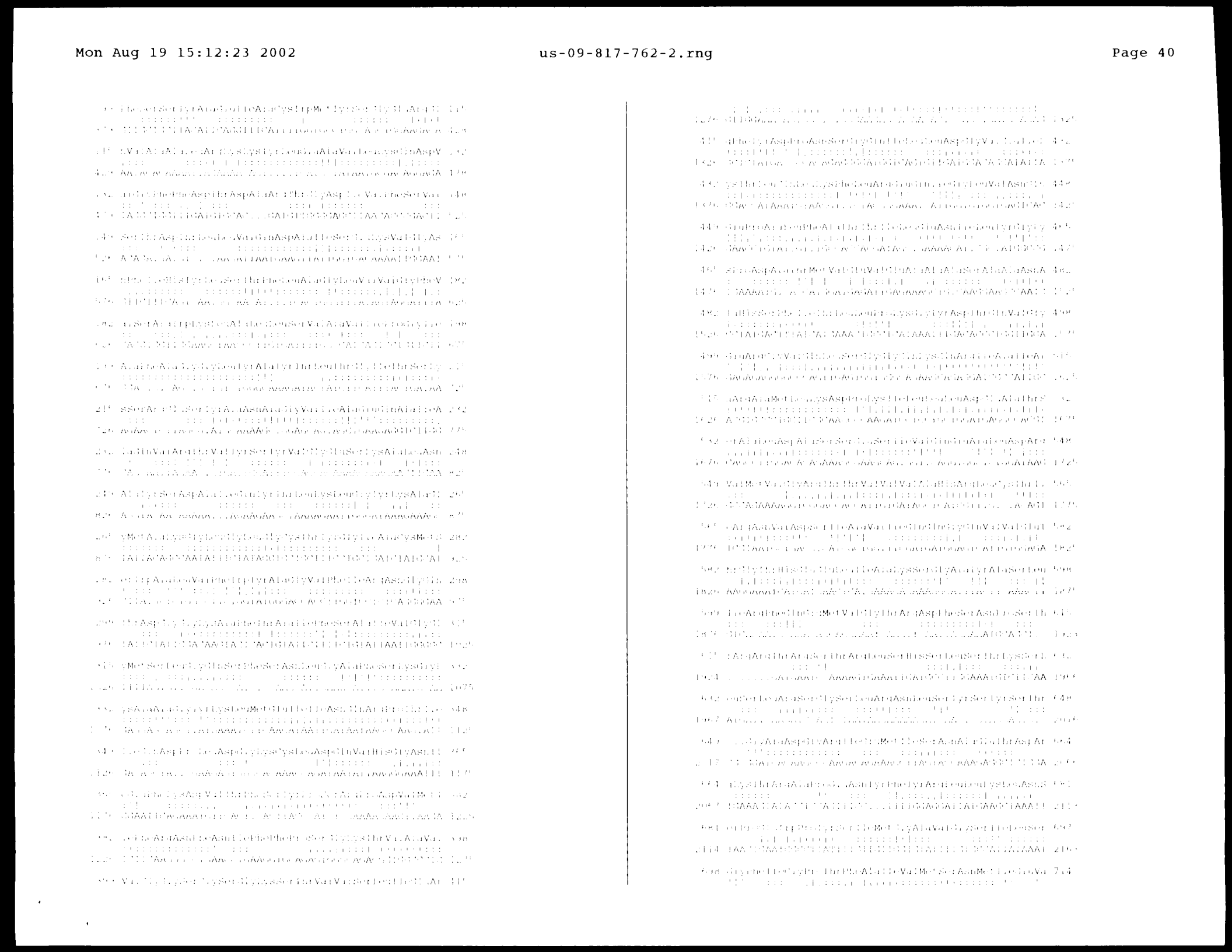


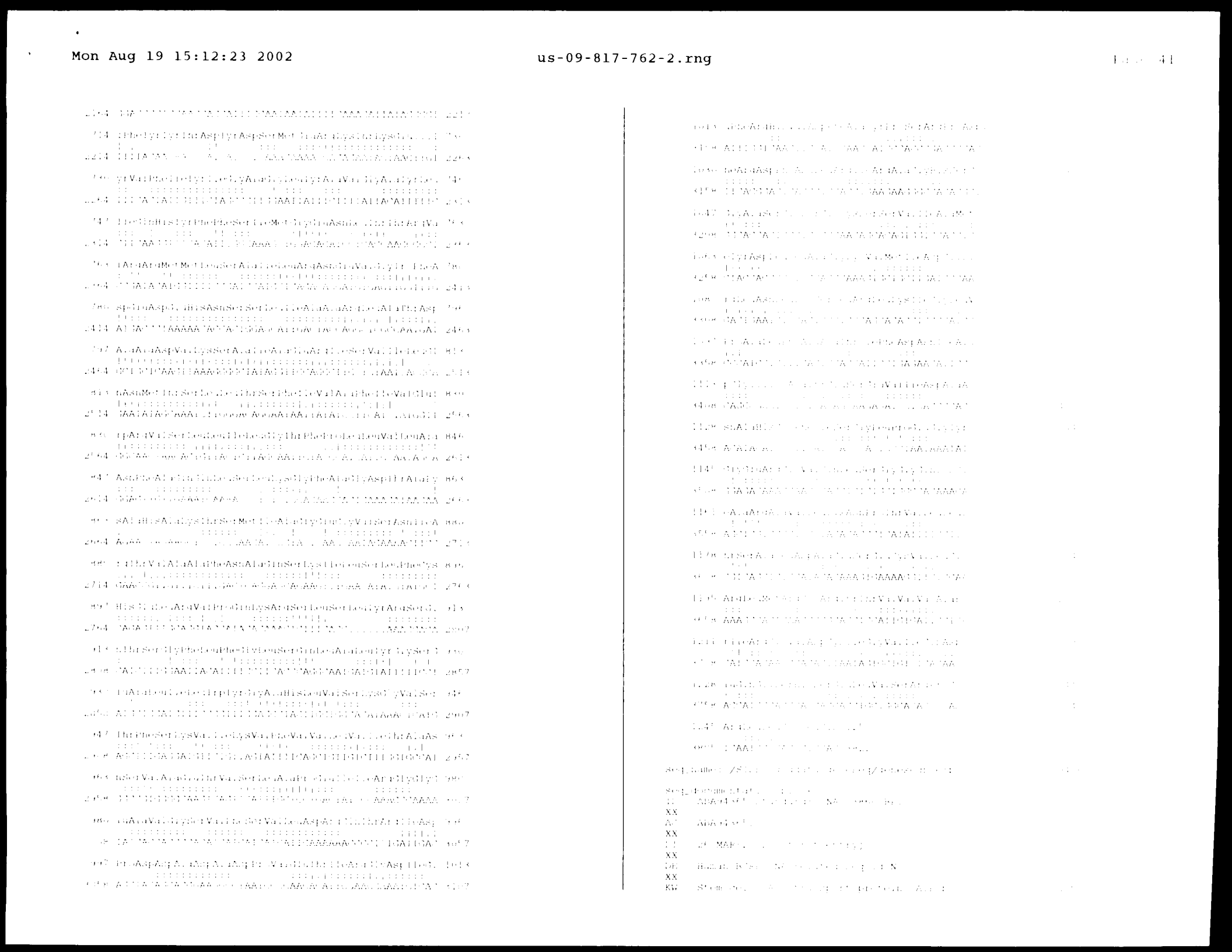


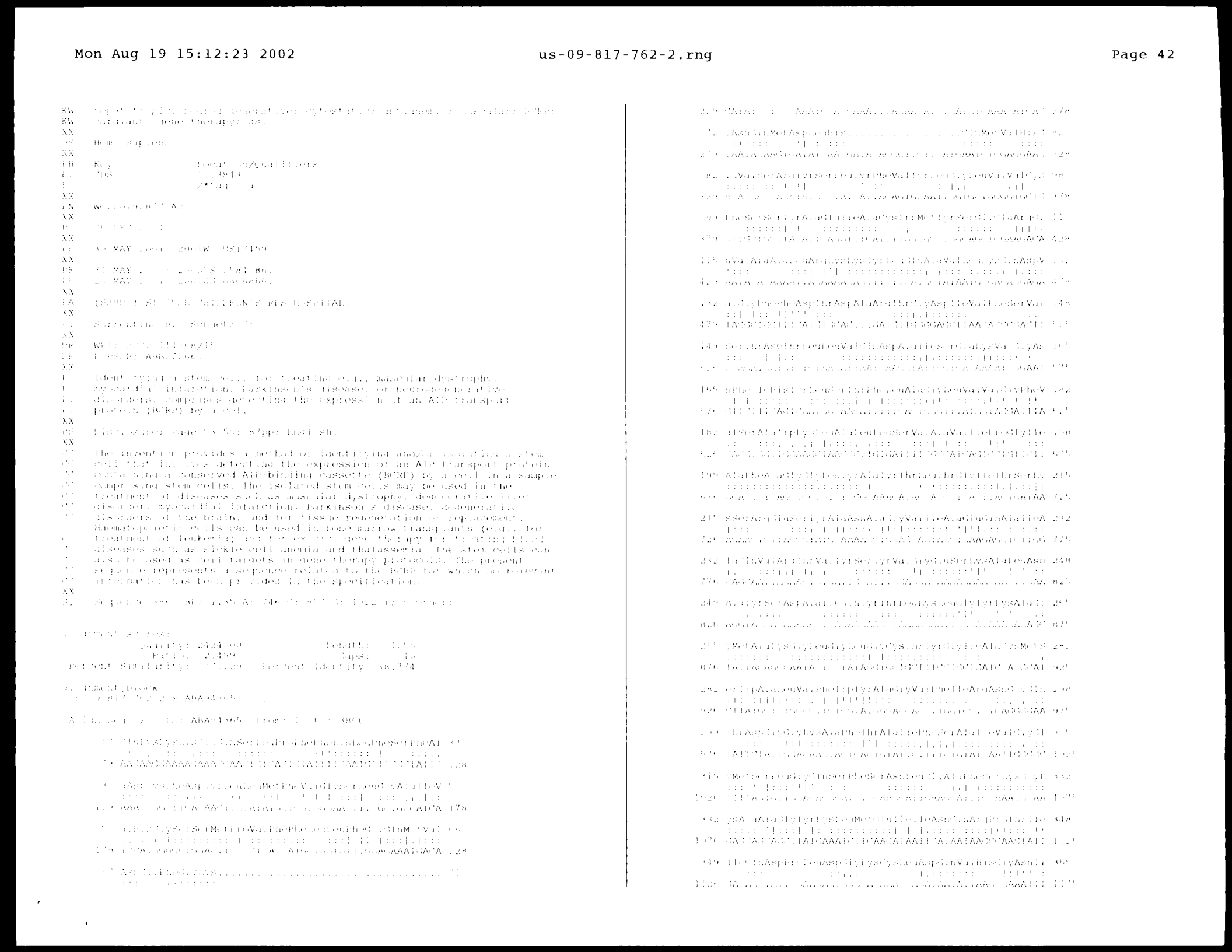


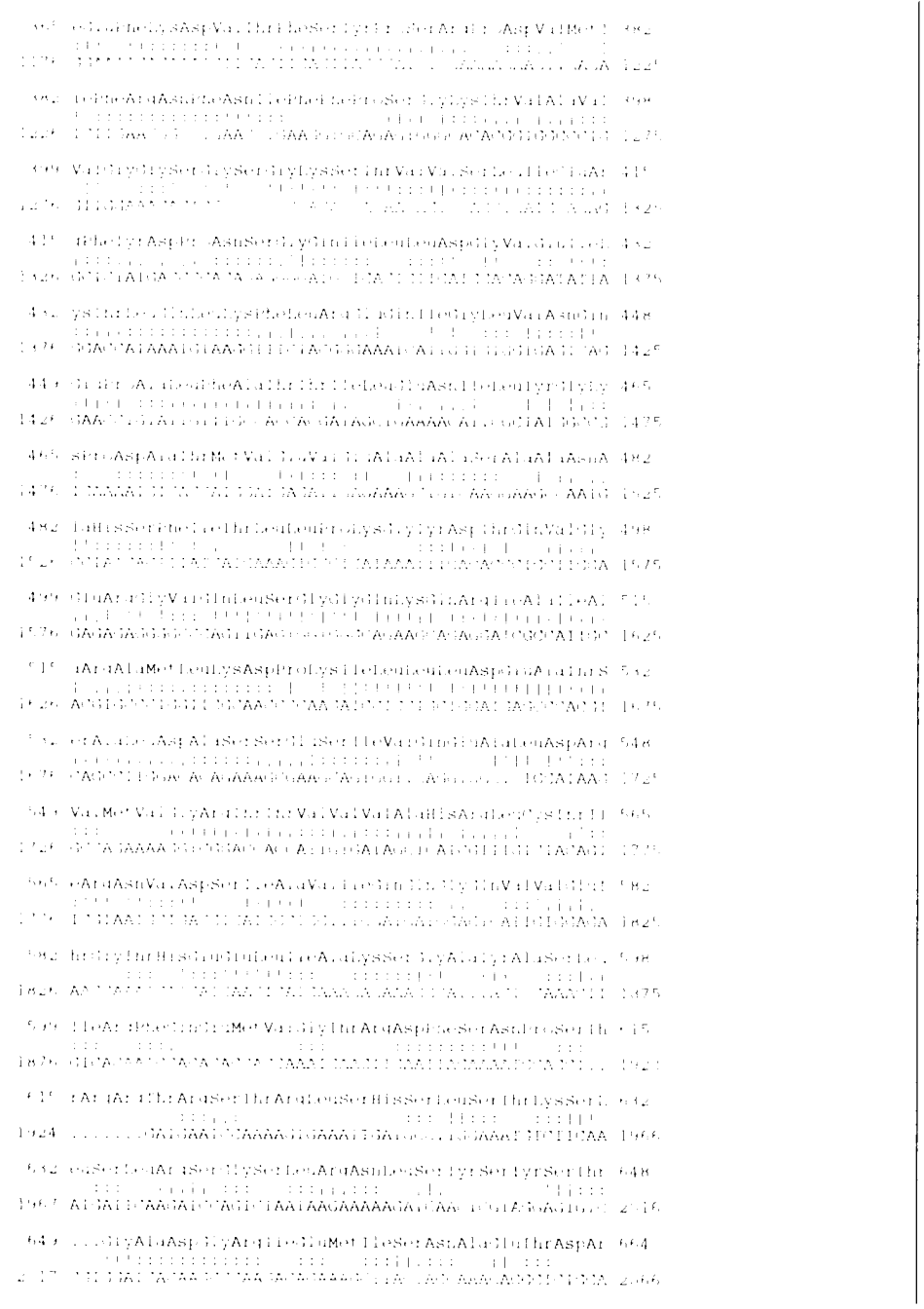


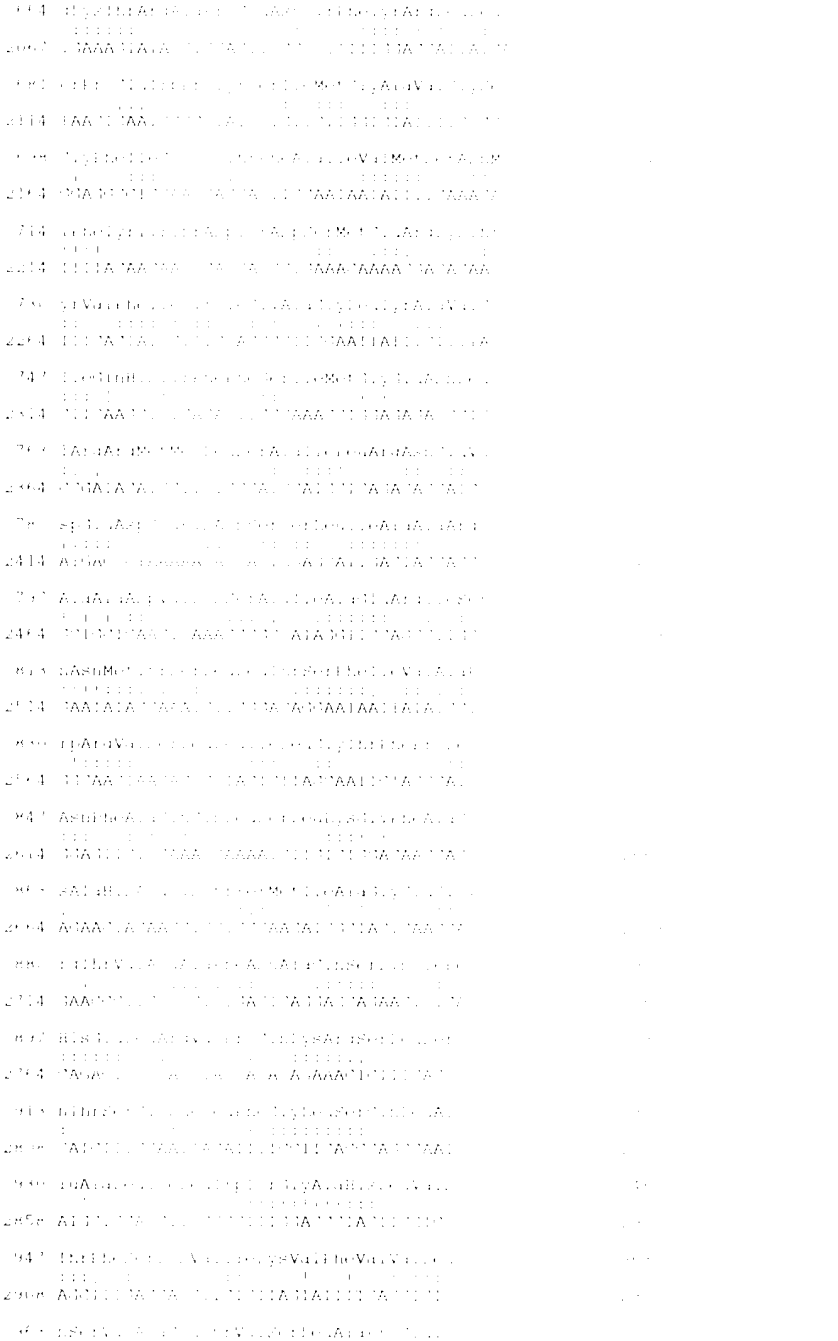


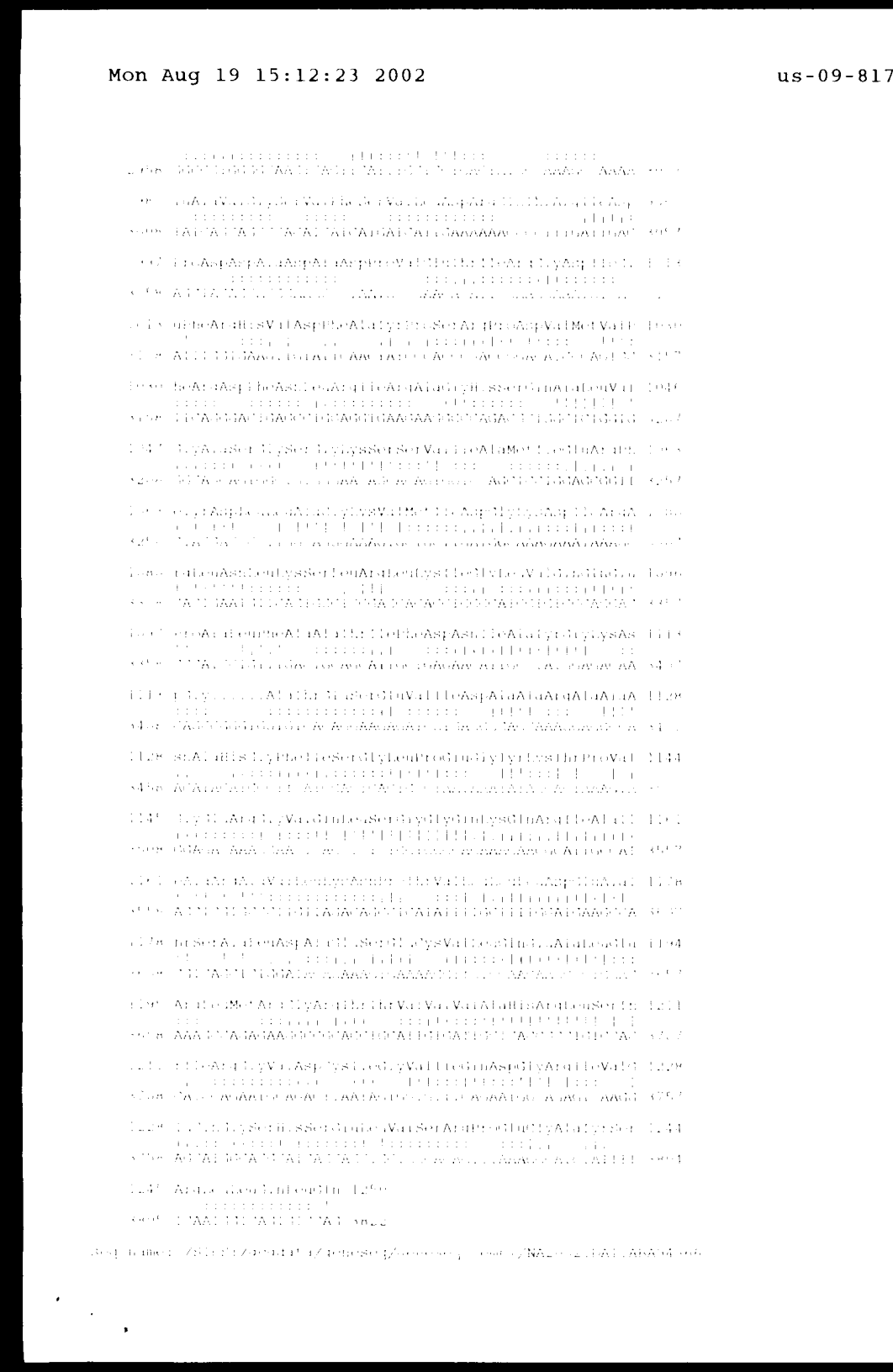


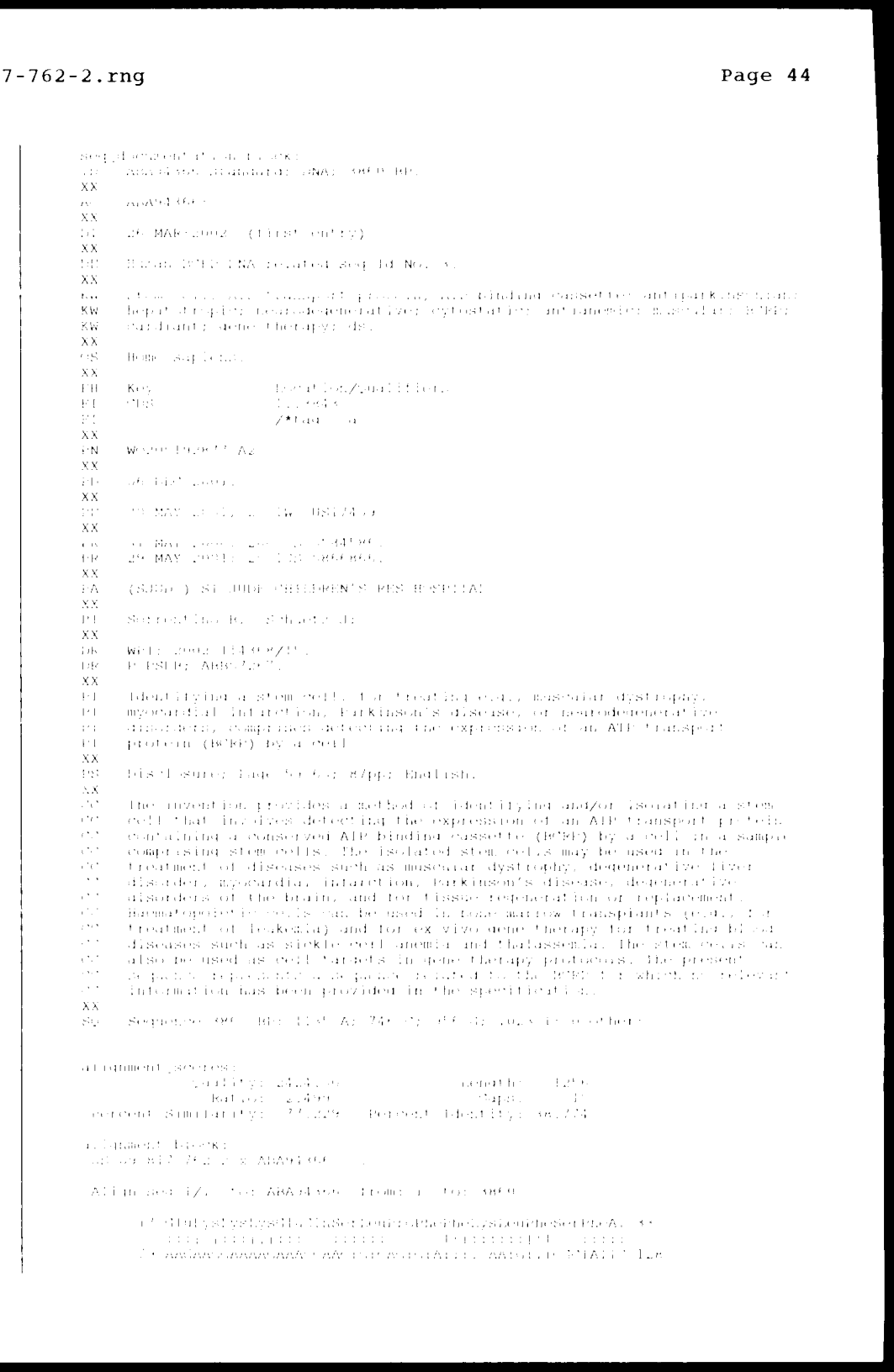


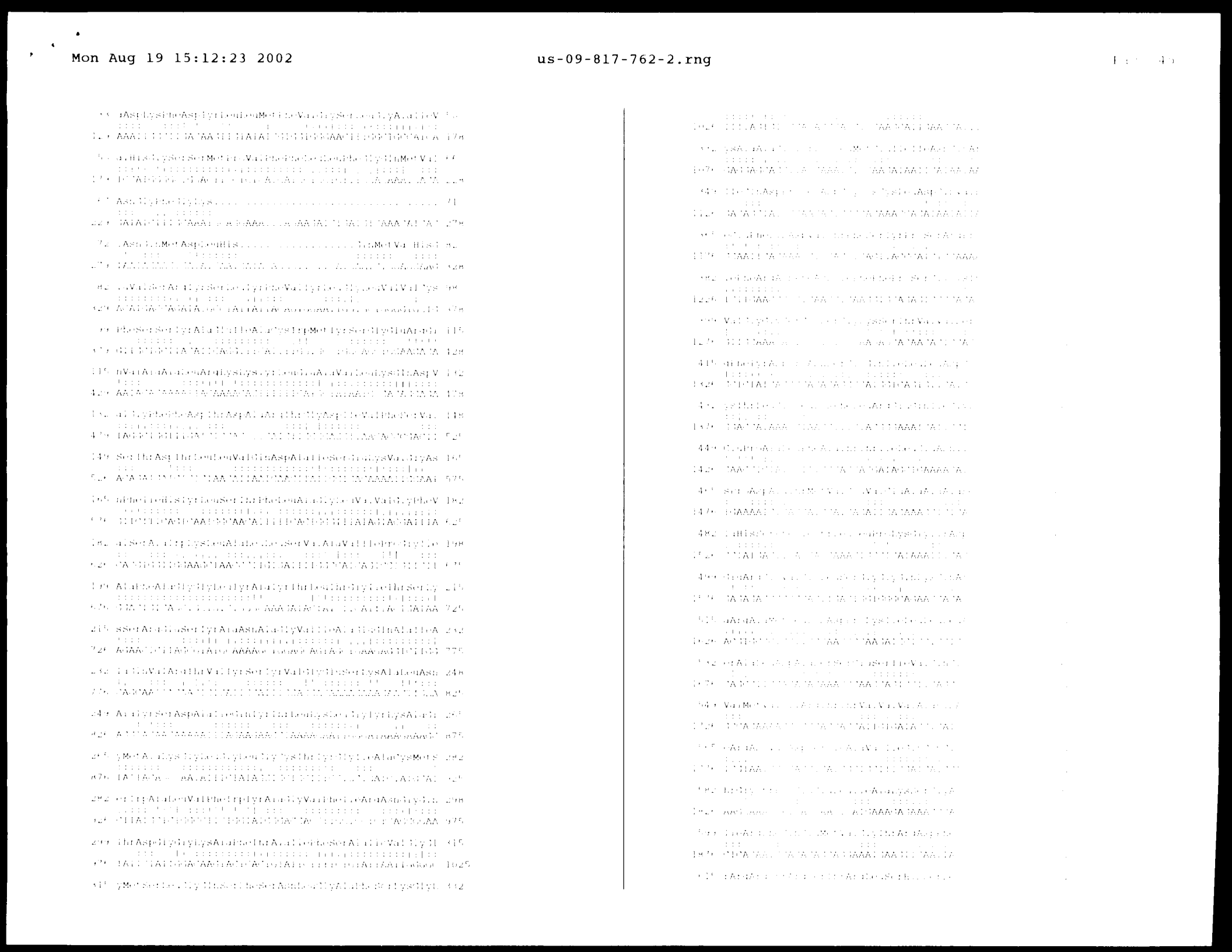


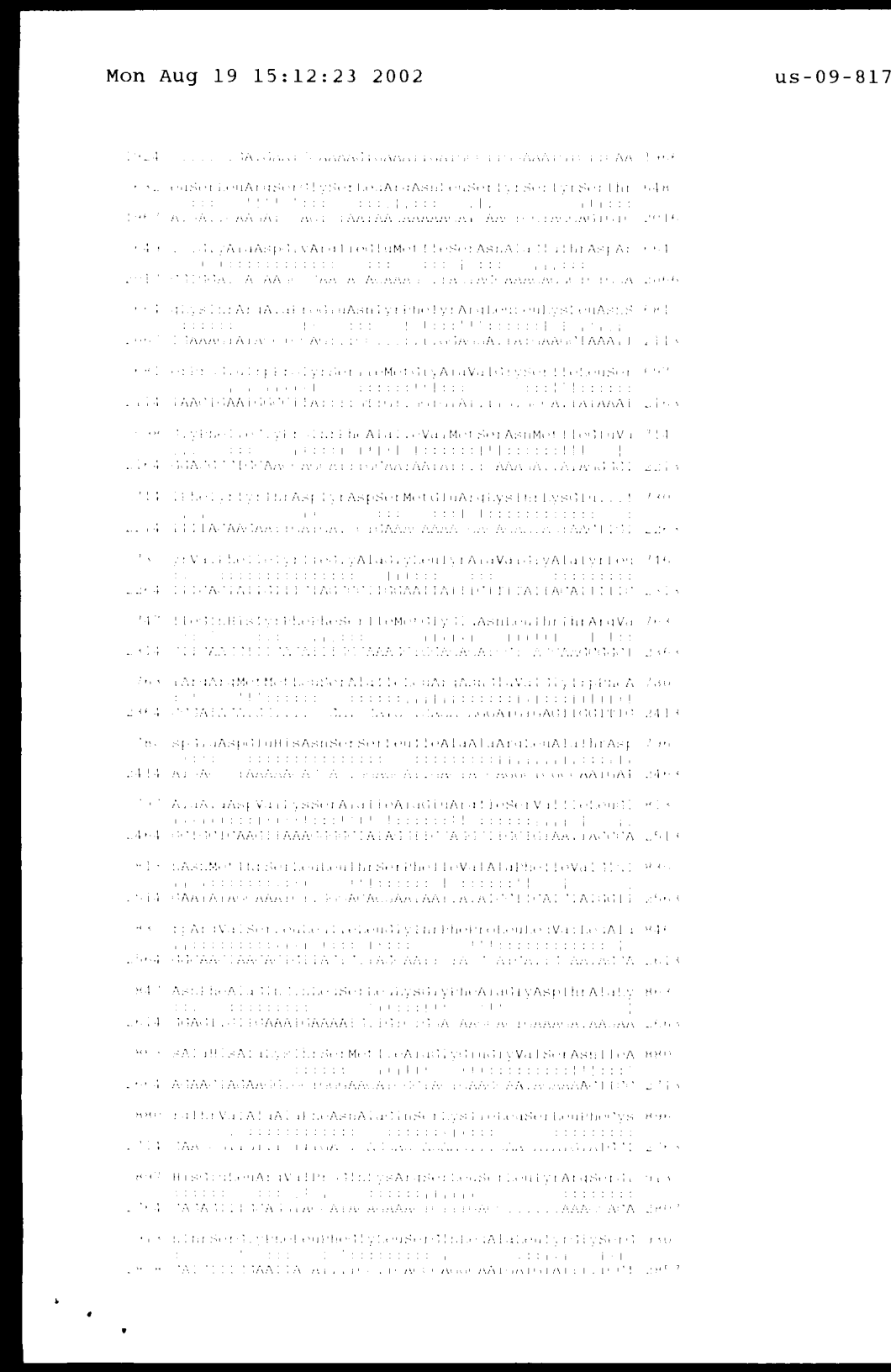


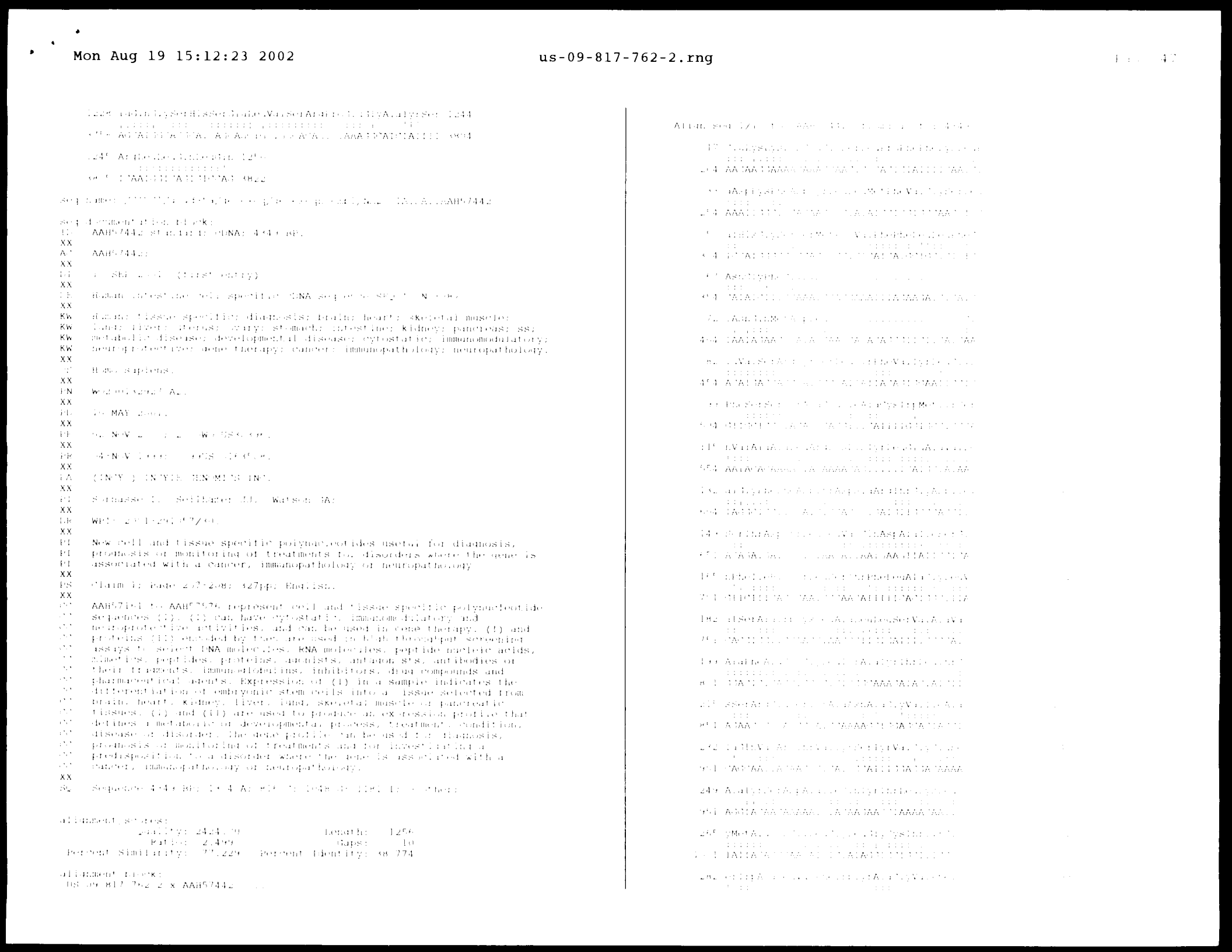


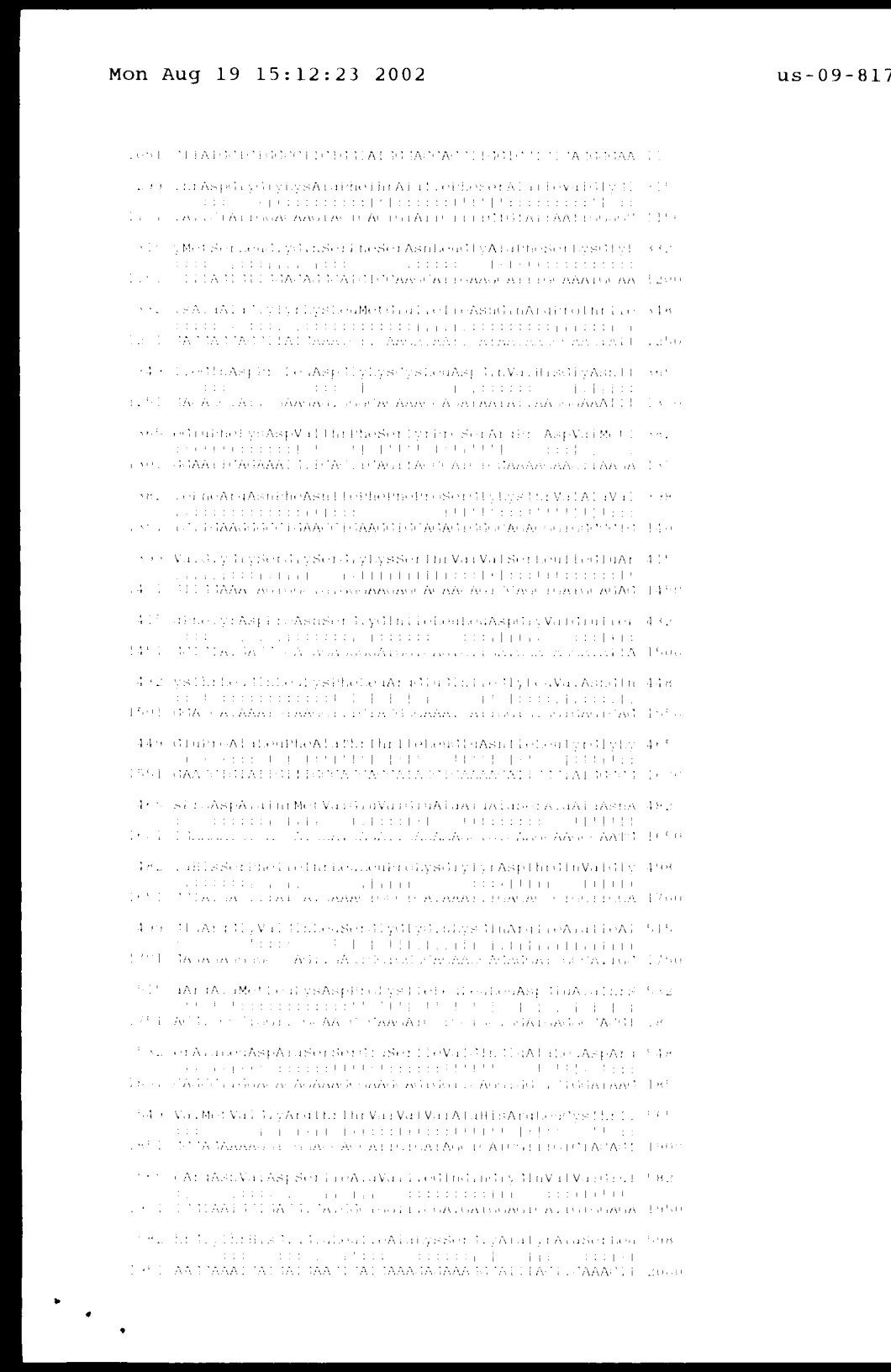


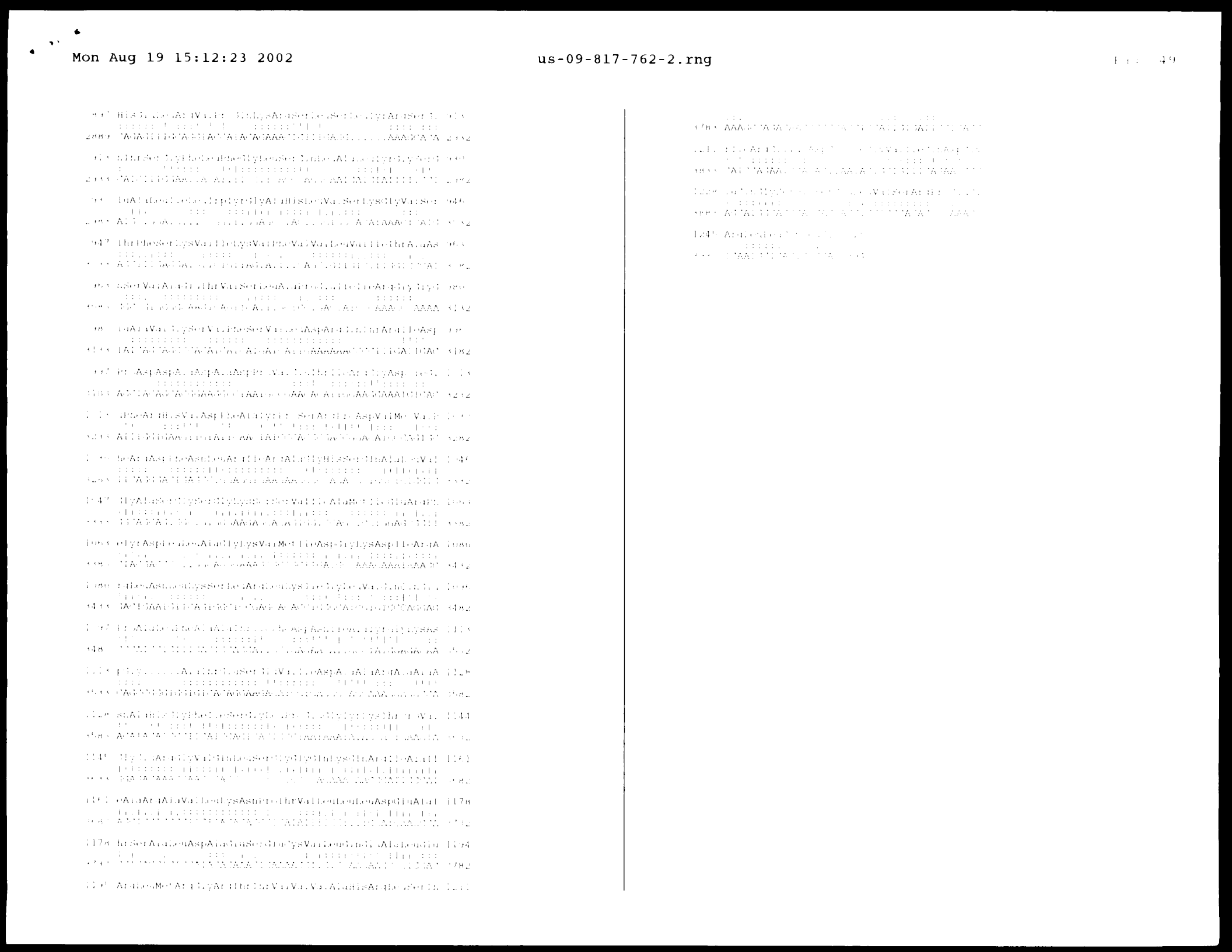


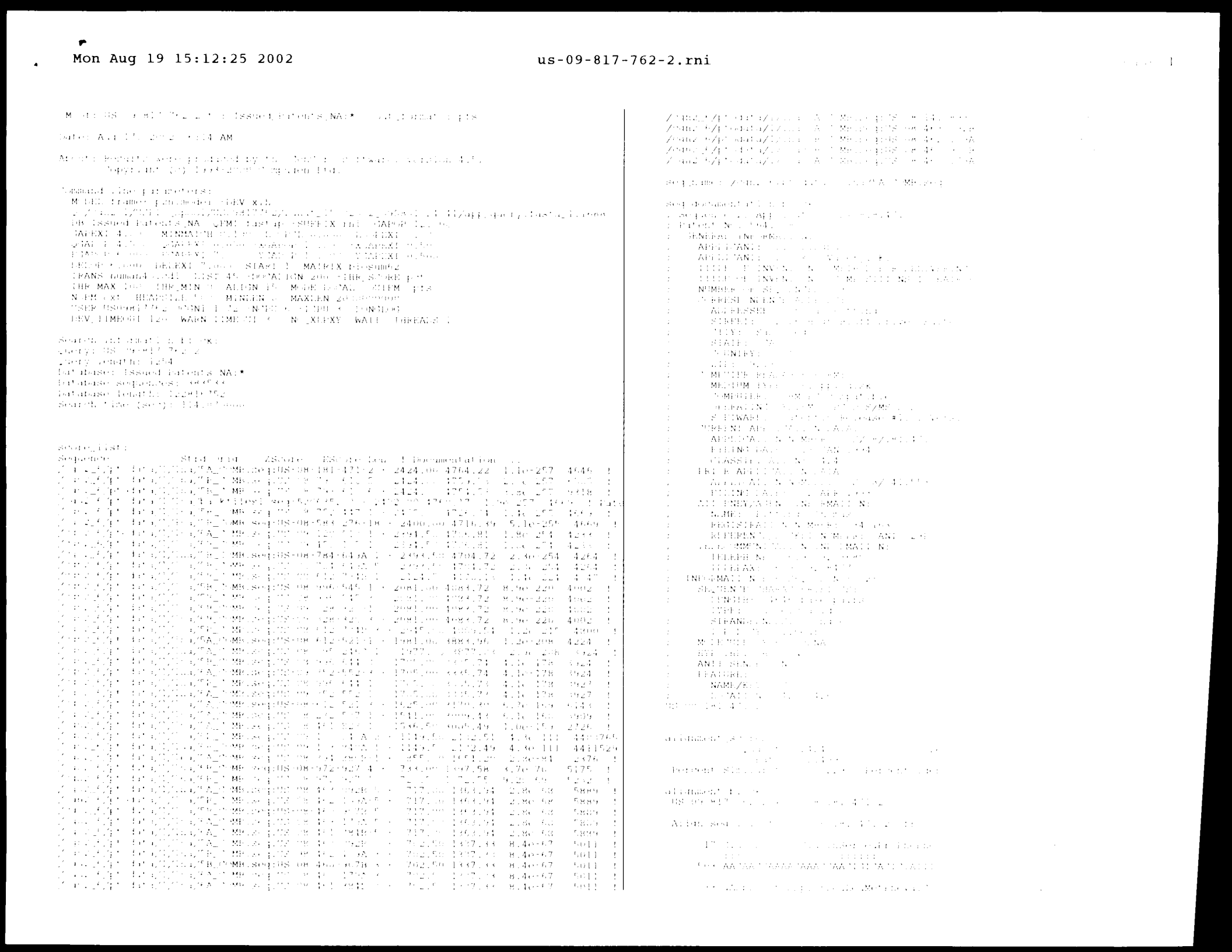


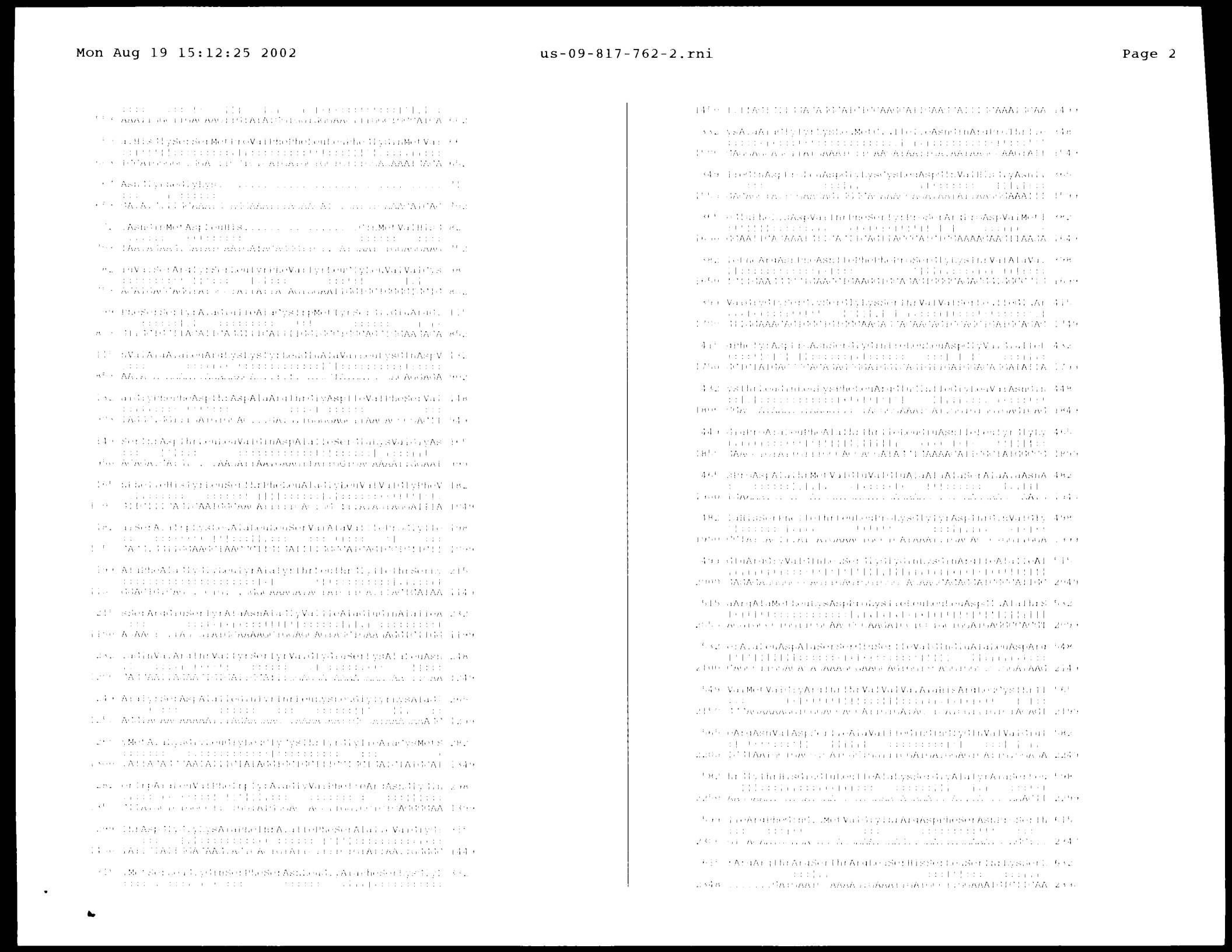


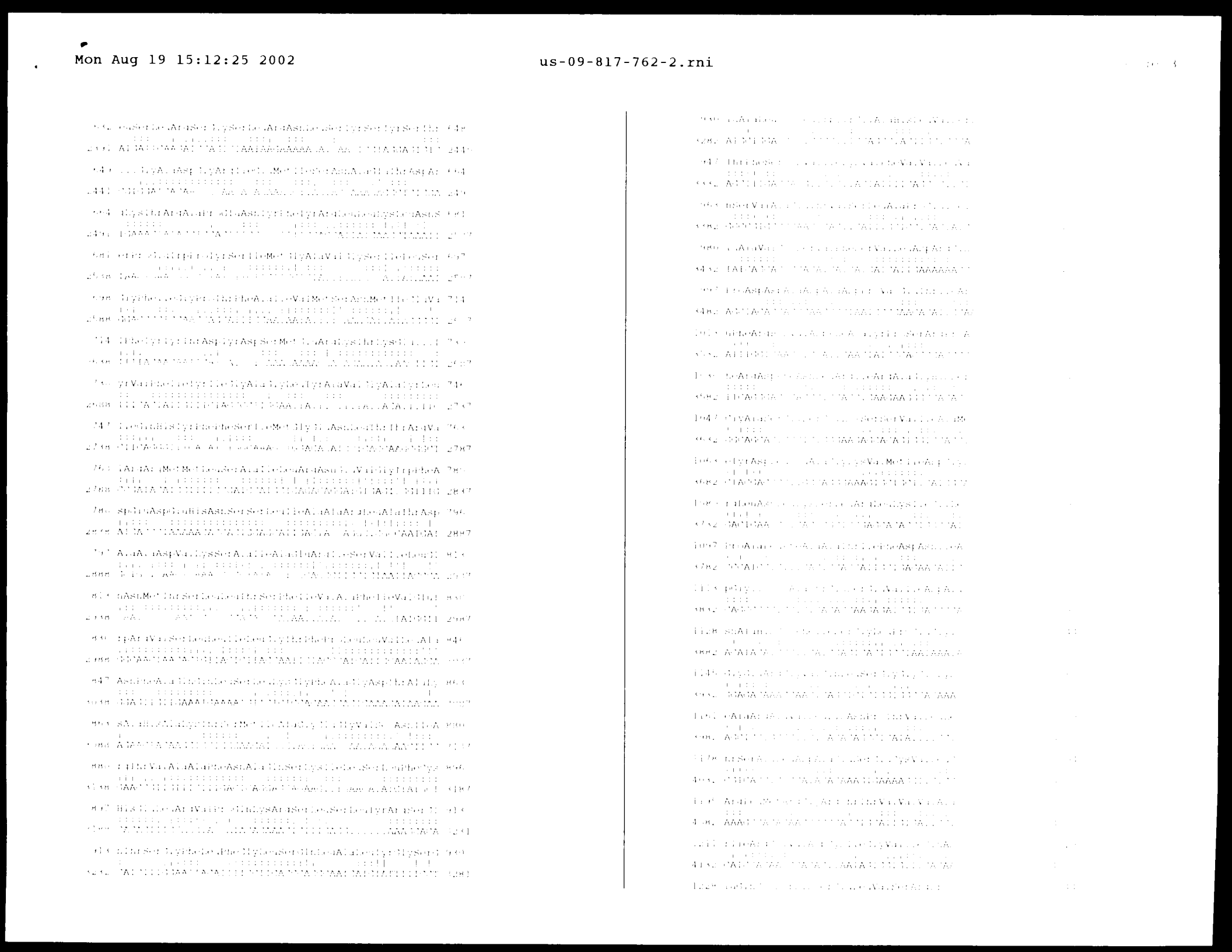


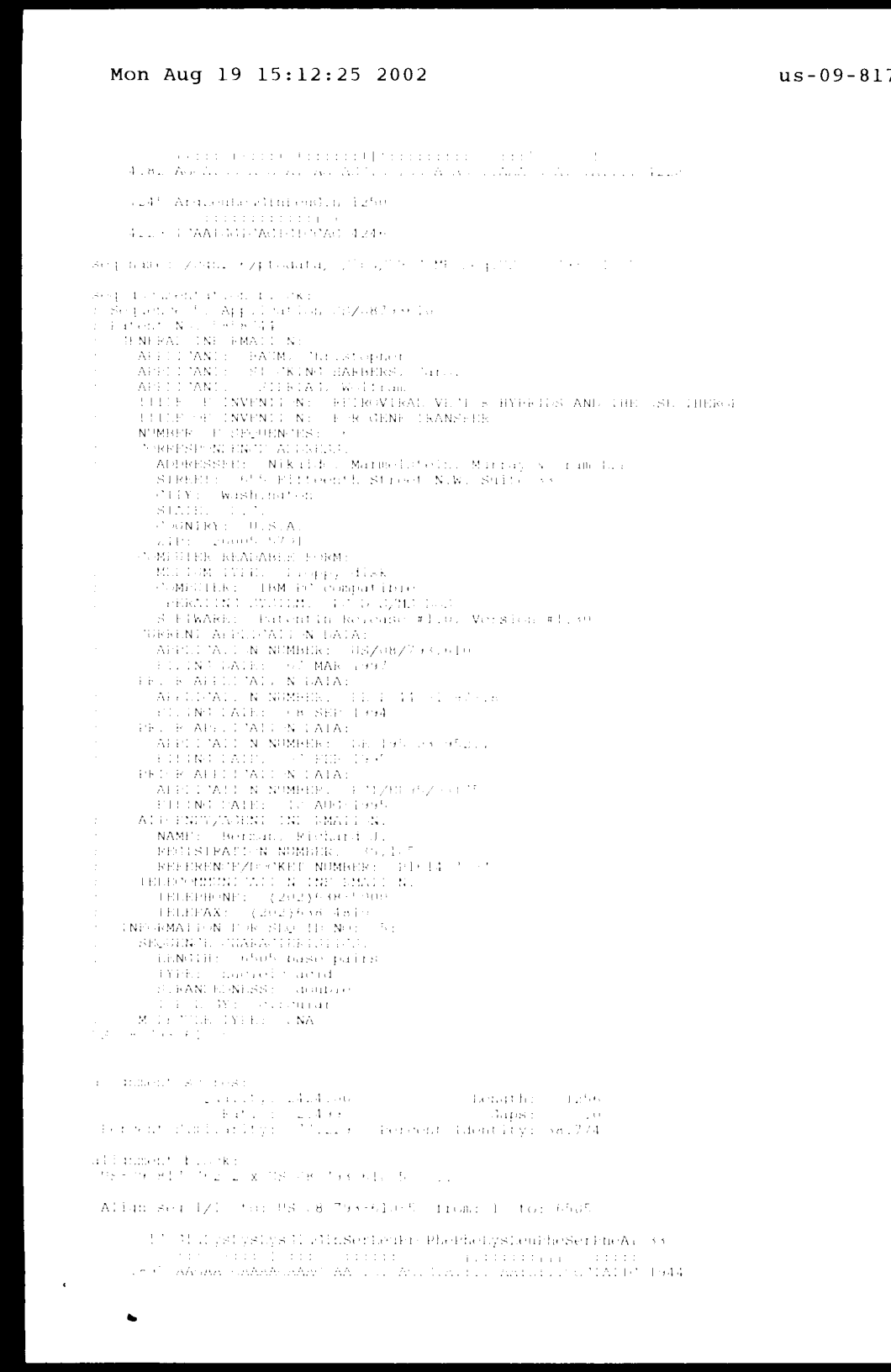


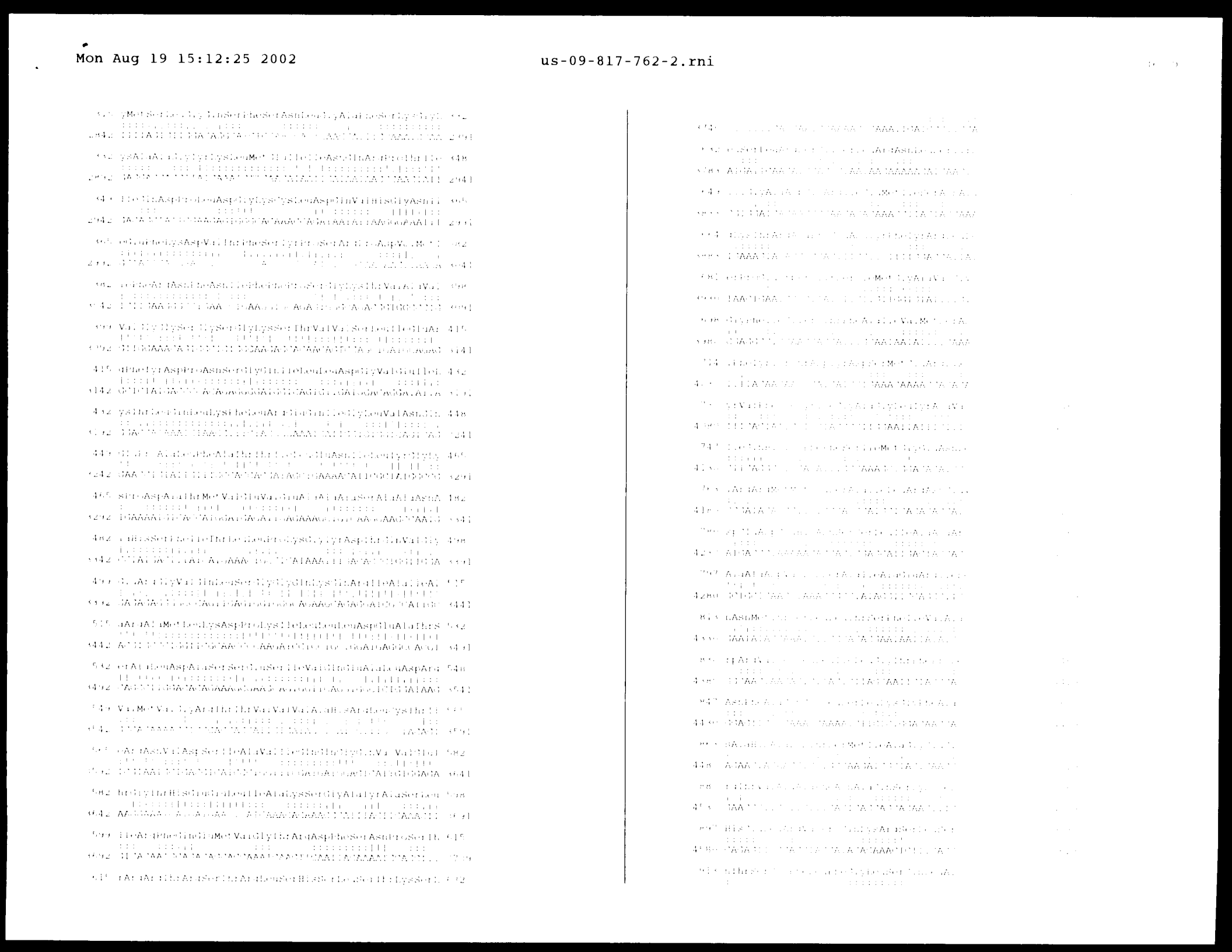


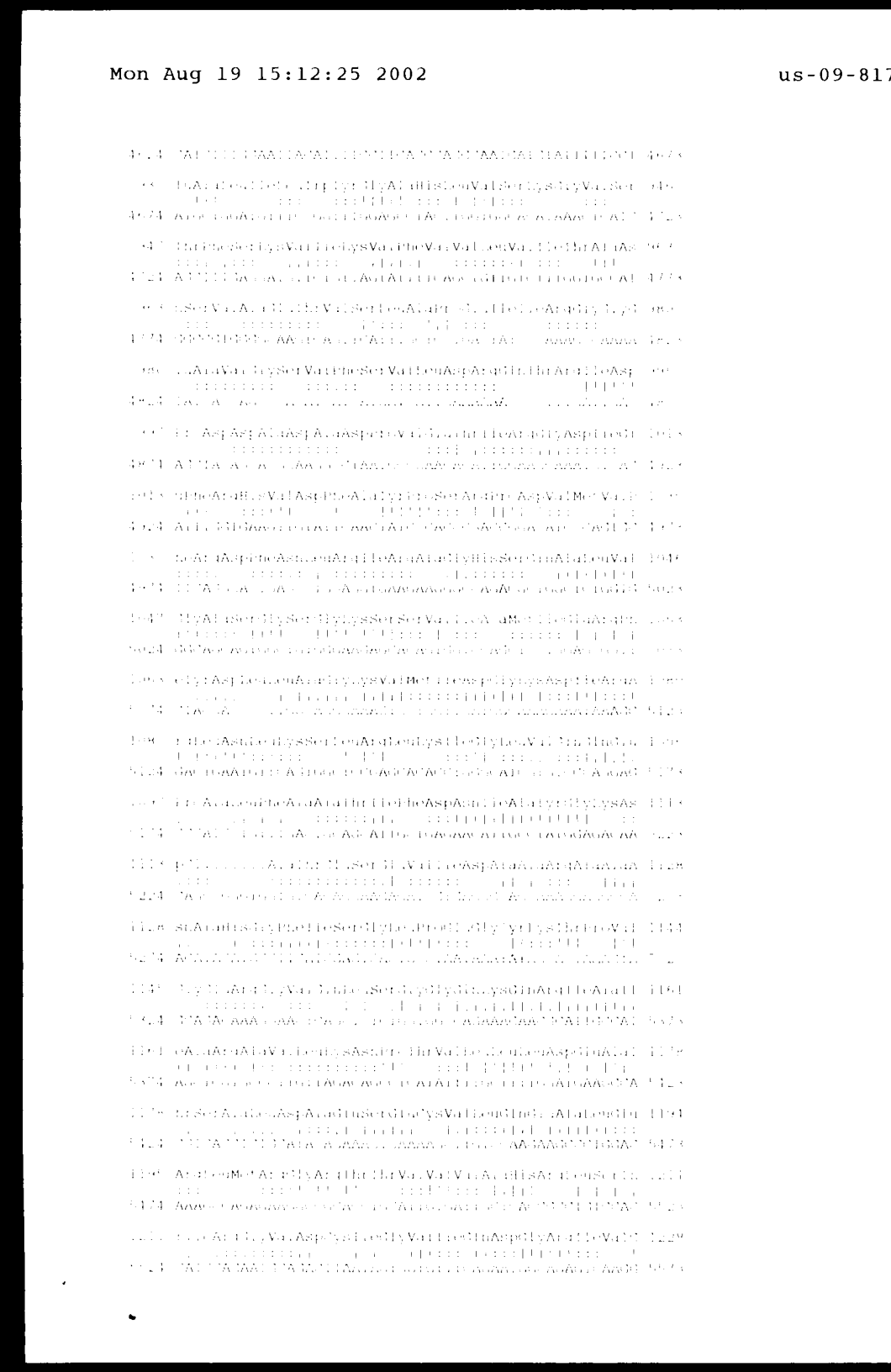


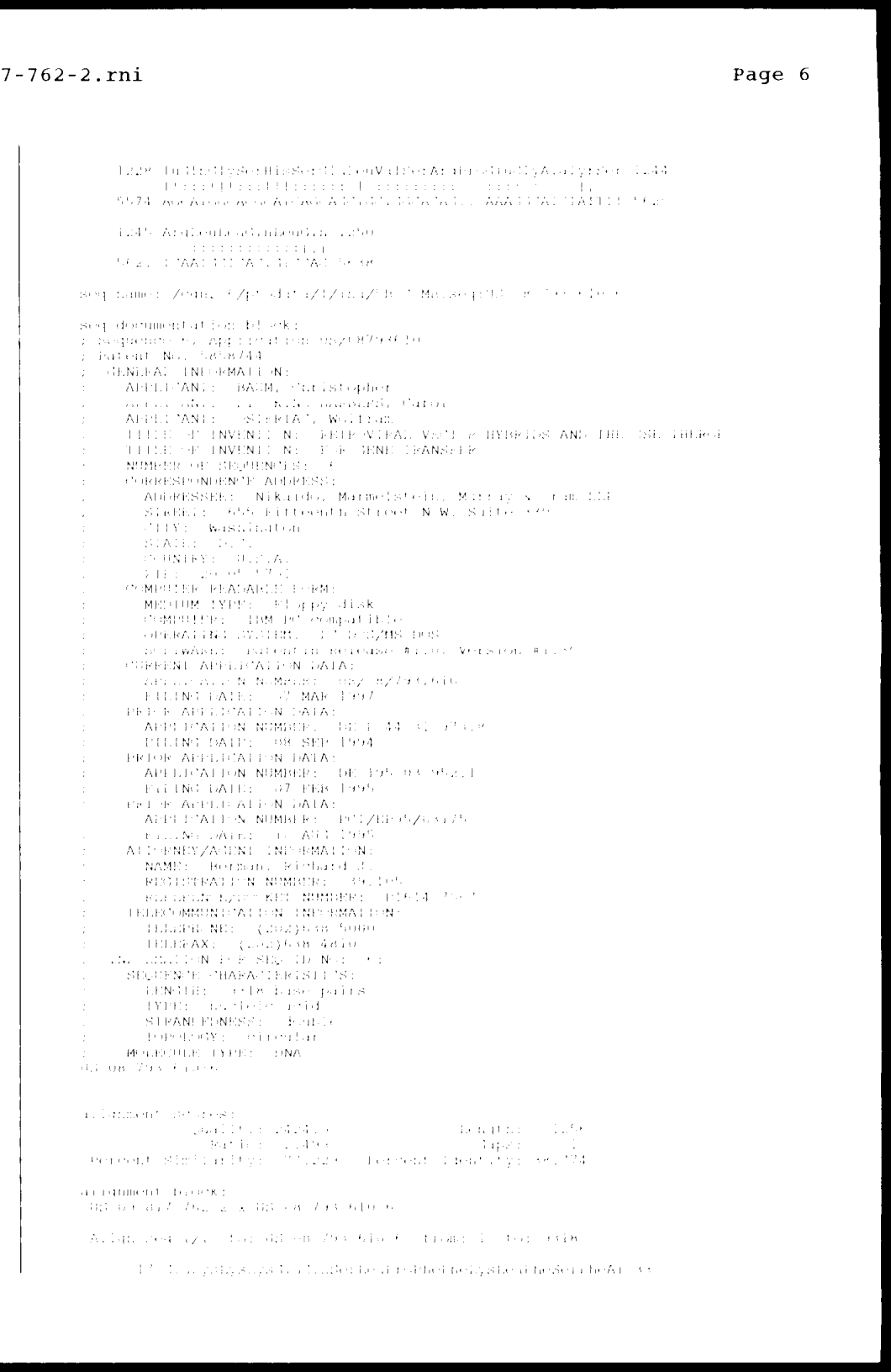


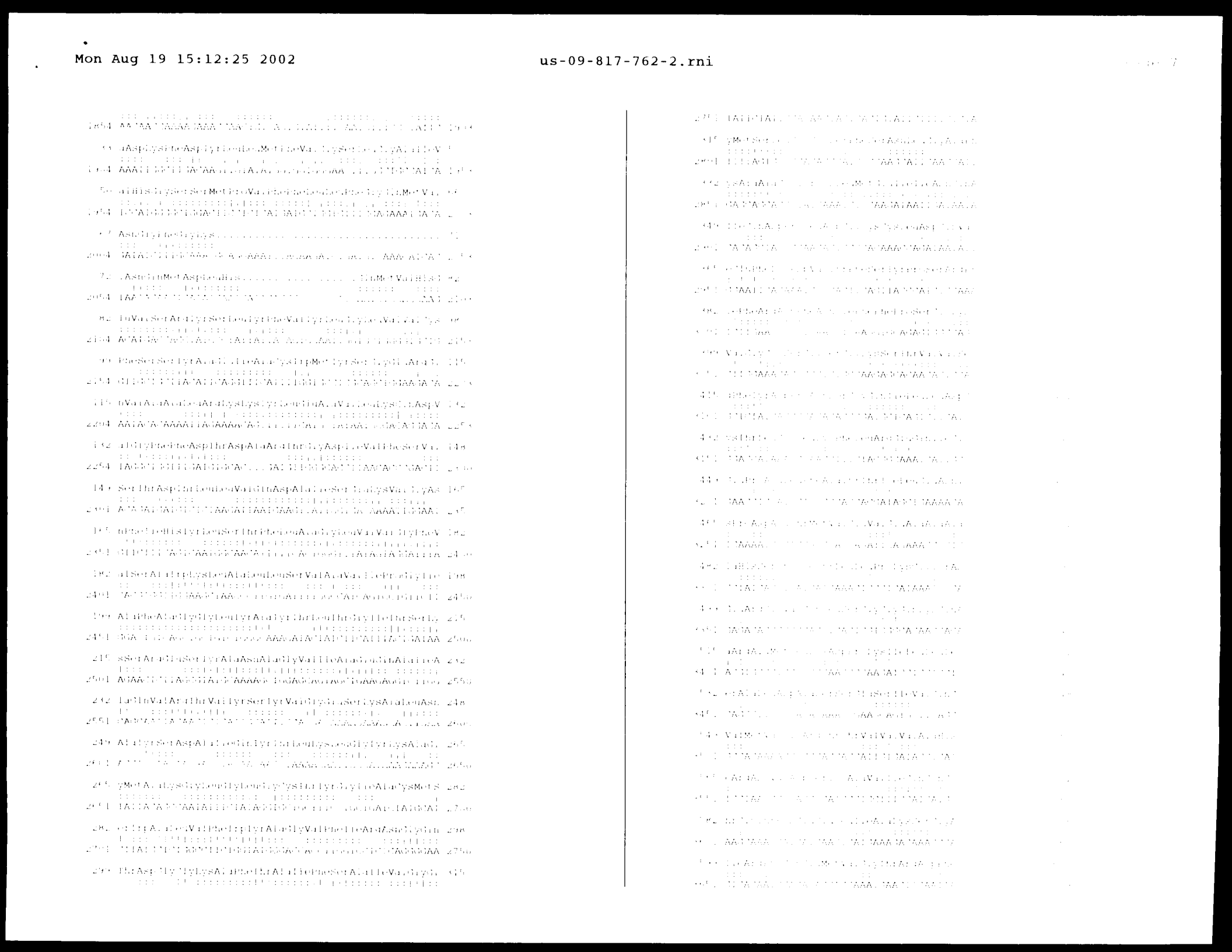


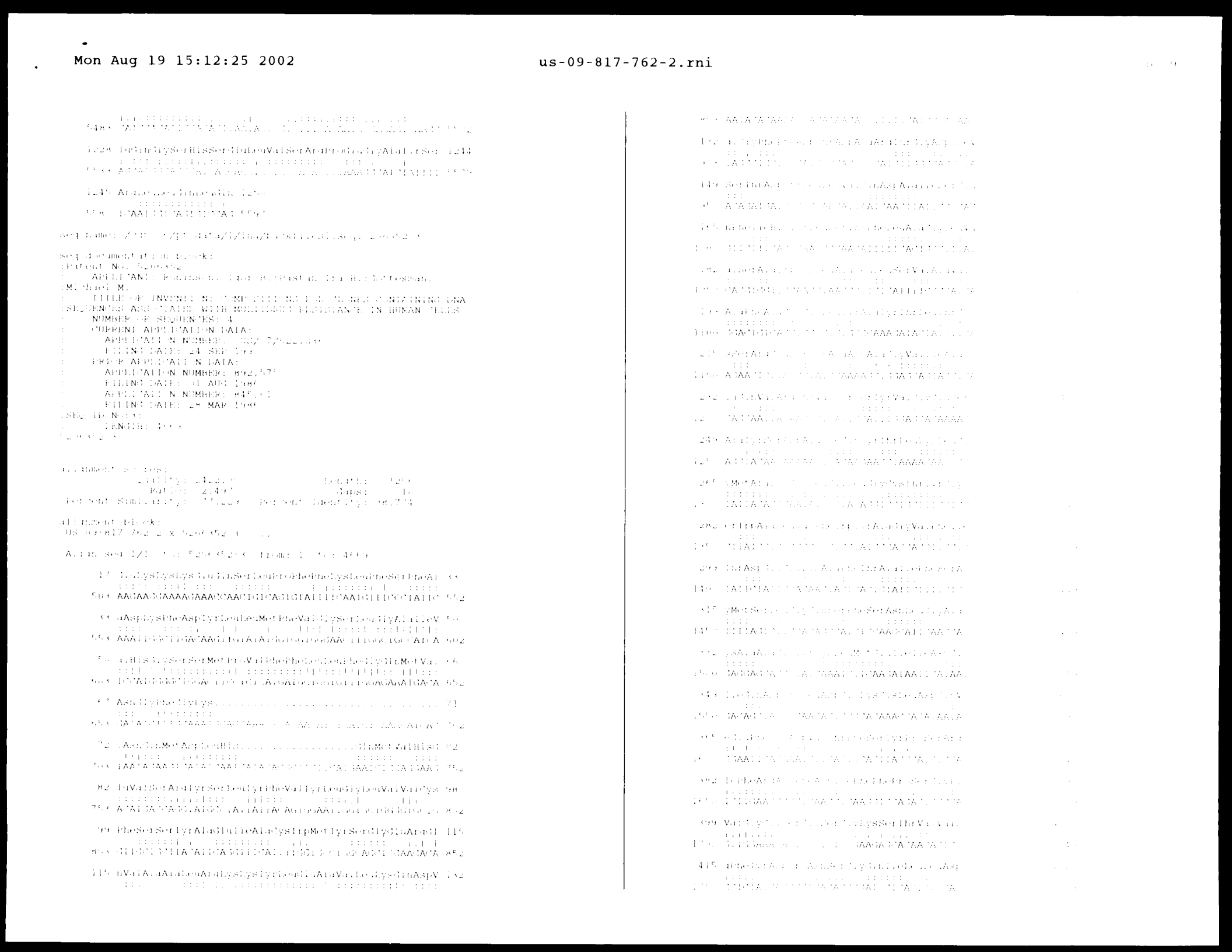


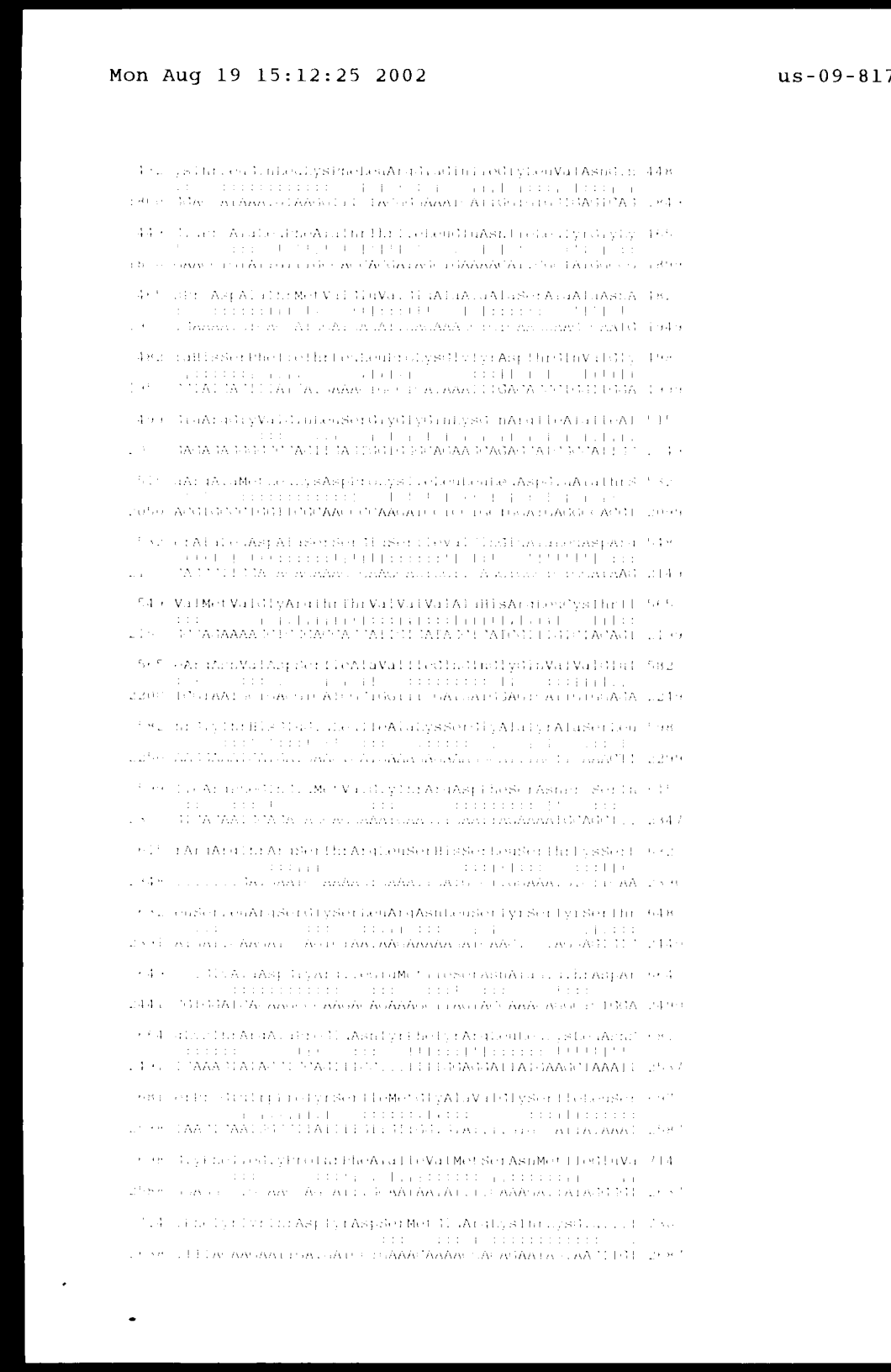


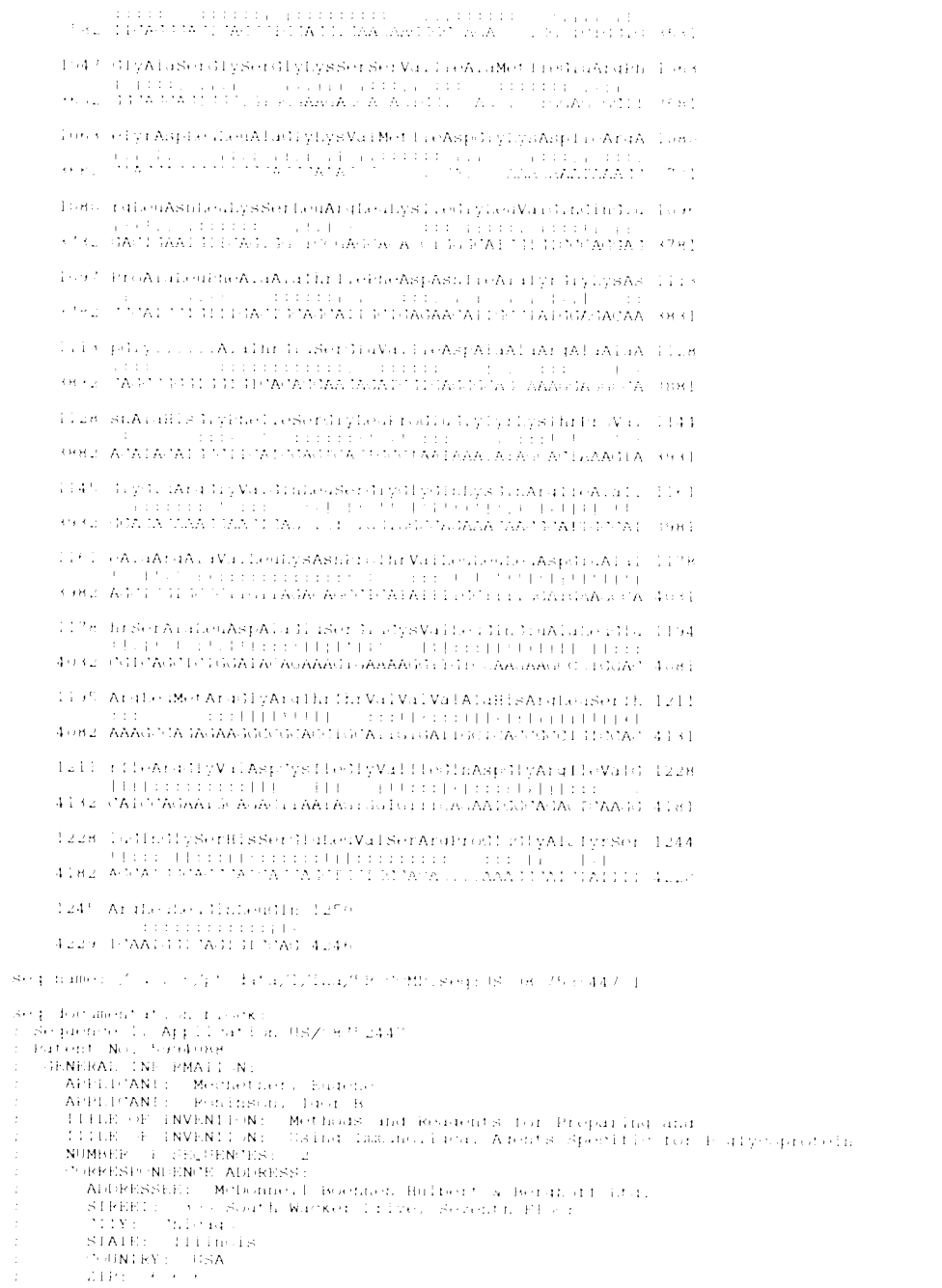


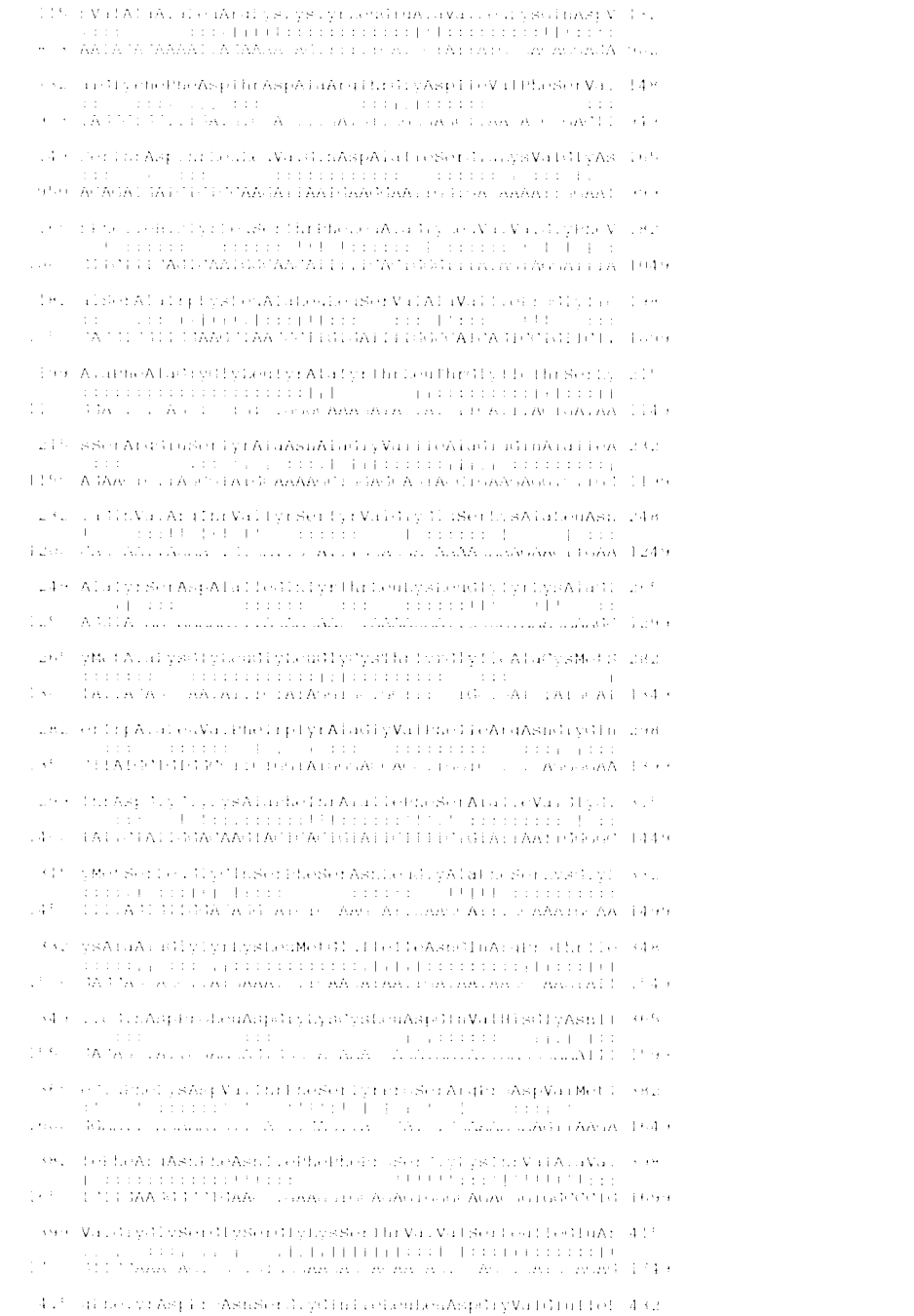


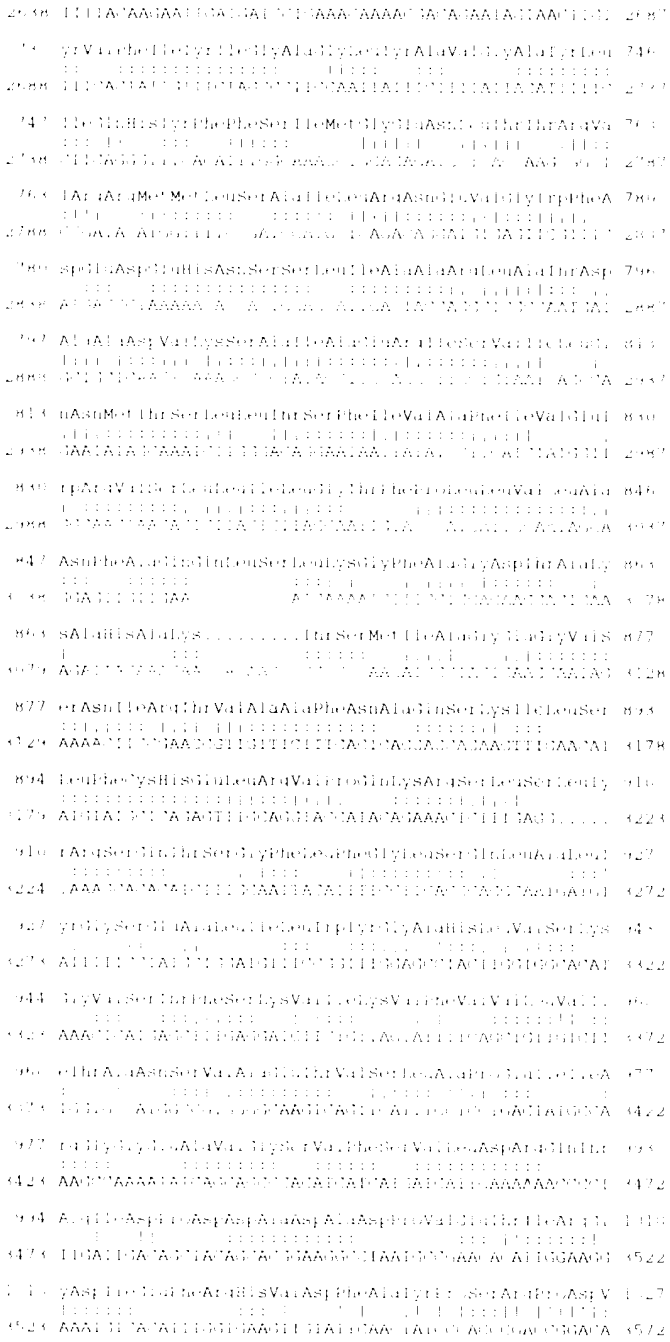


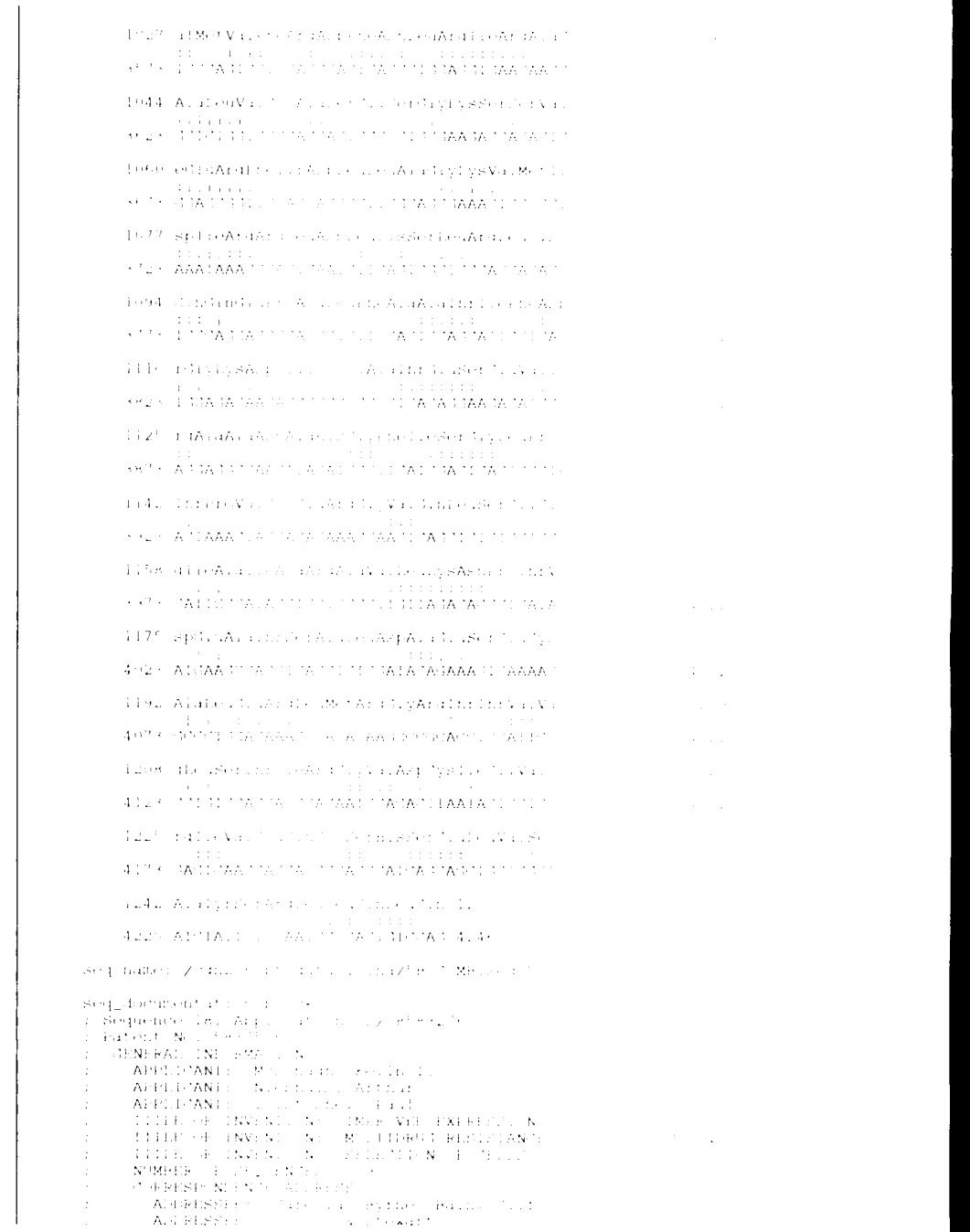


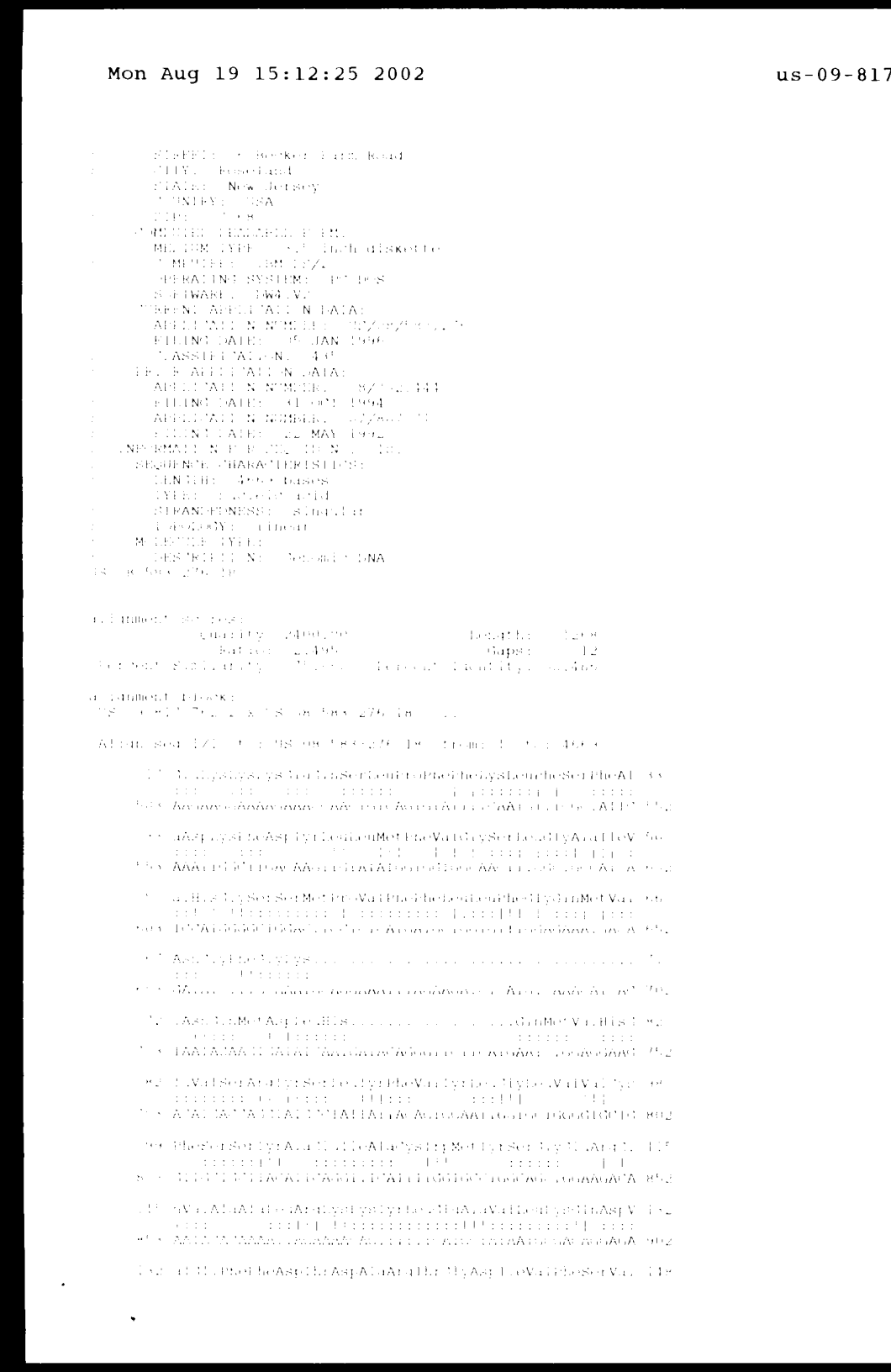


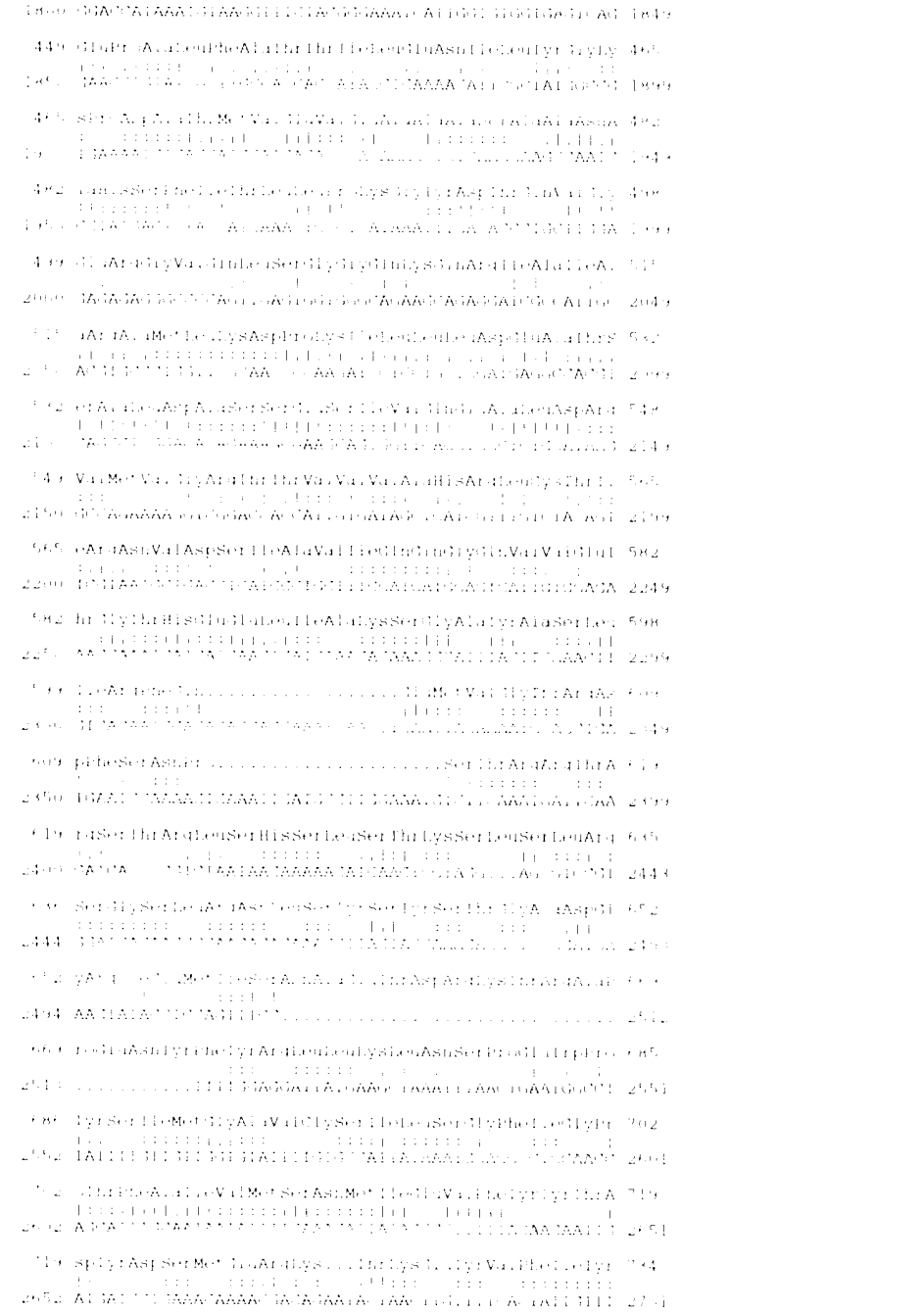


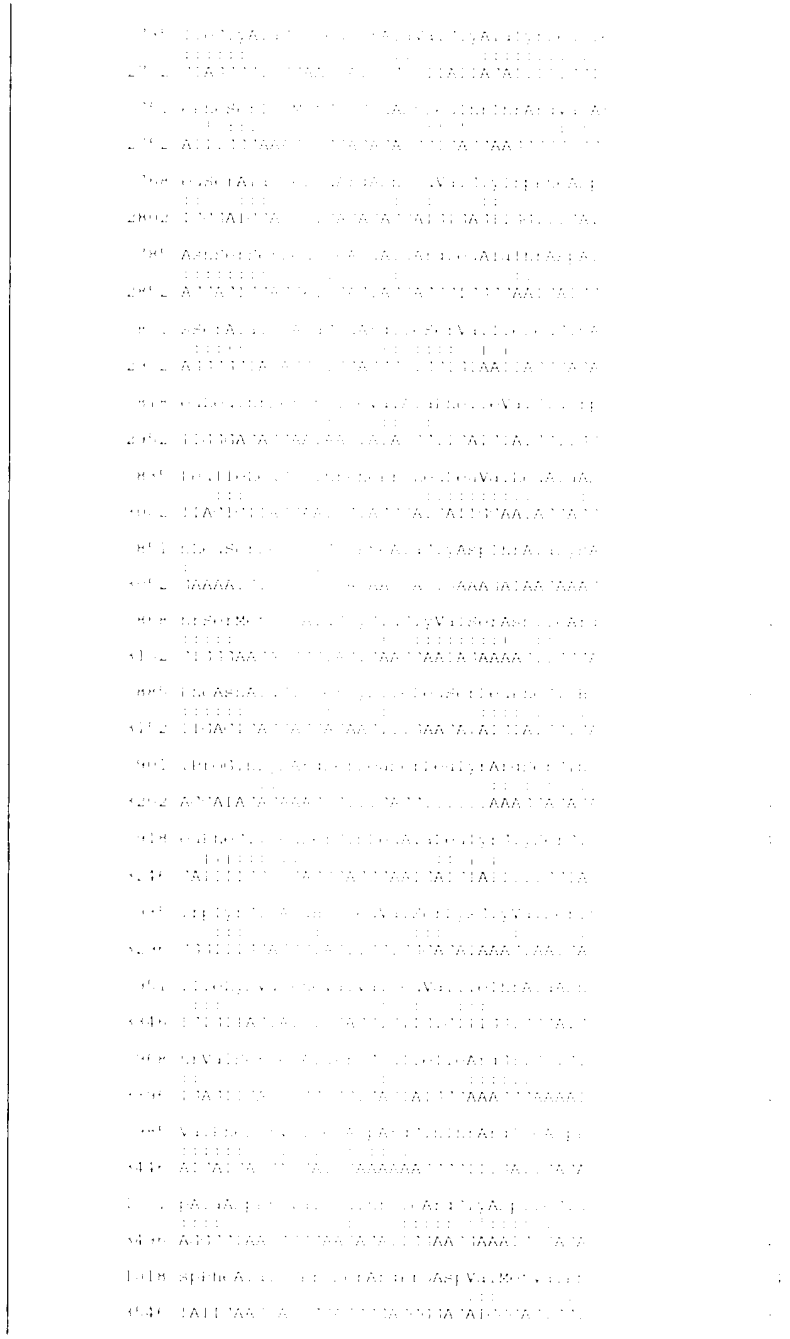












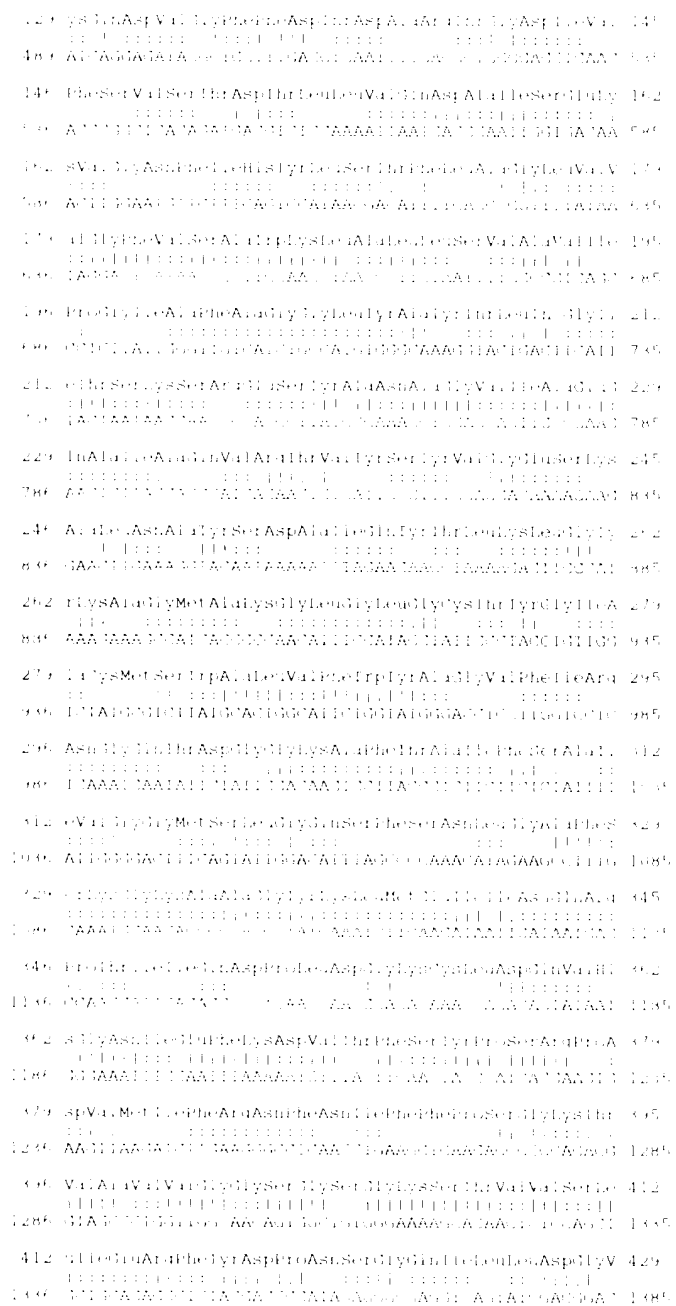
[illegible]

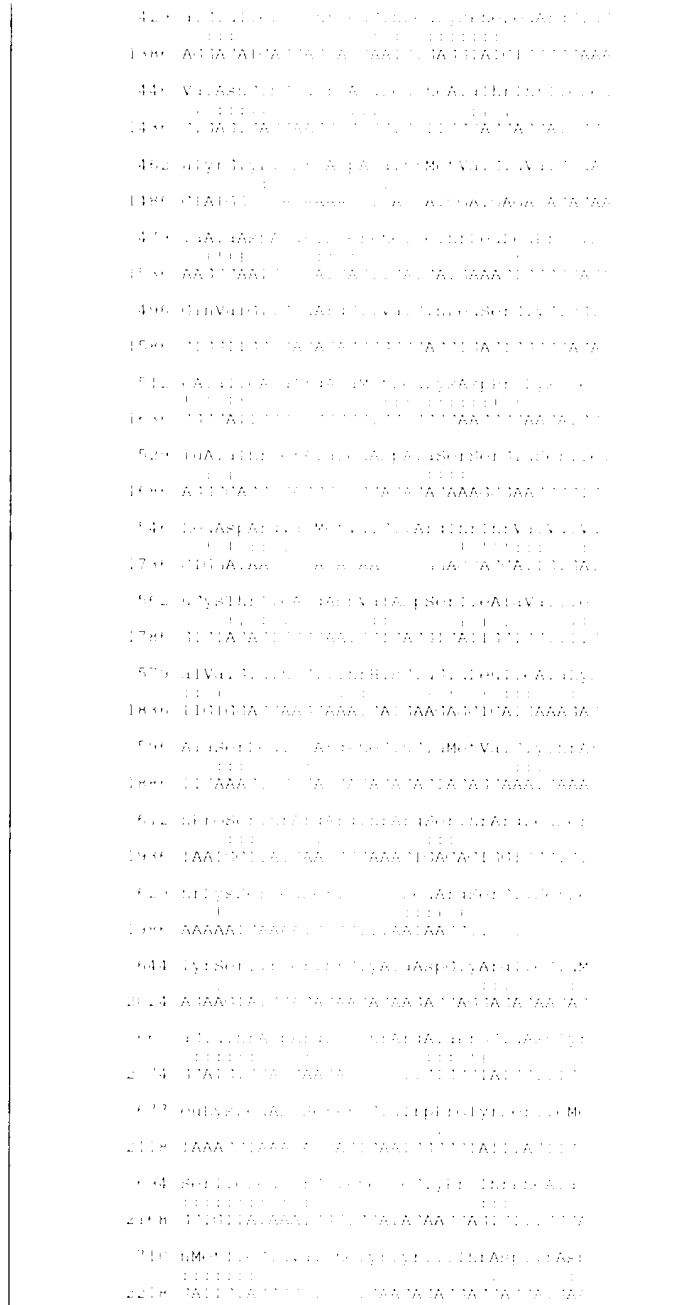
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY:
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,513
; FILING DATE: [illegible]
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: [illegible]
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: [illegible]
; TELEFAX: [illegible]
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA

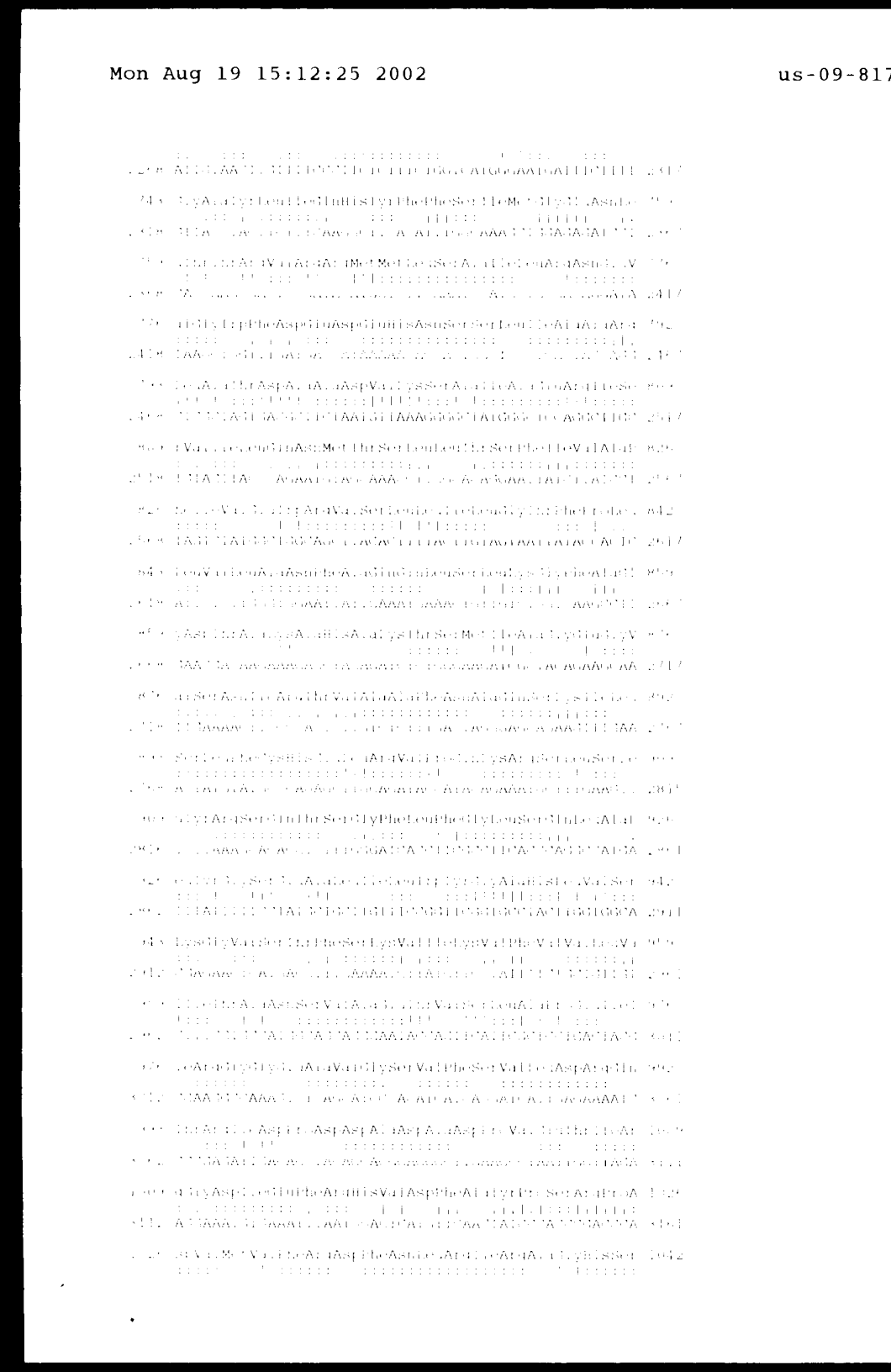
Alignment Scores:
Quality: [illegible] Length: [illegible]
Ratio: [illegible] Gaps: [illegible]
Percent Similarity: [illegible] Percent Identity: [illegible]

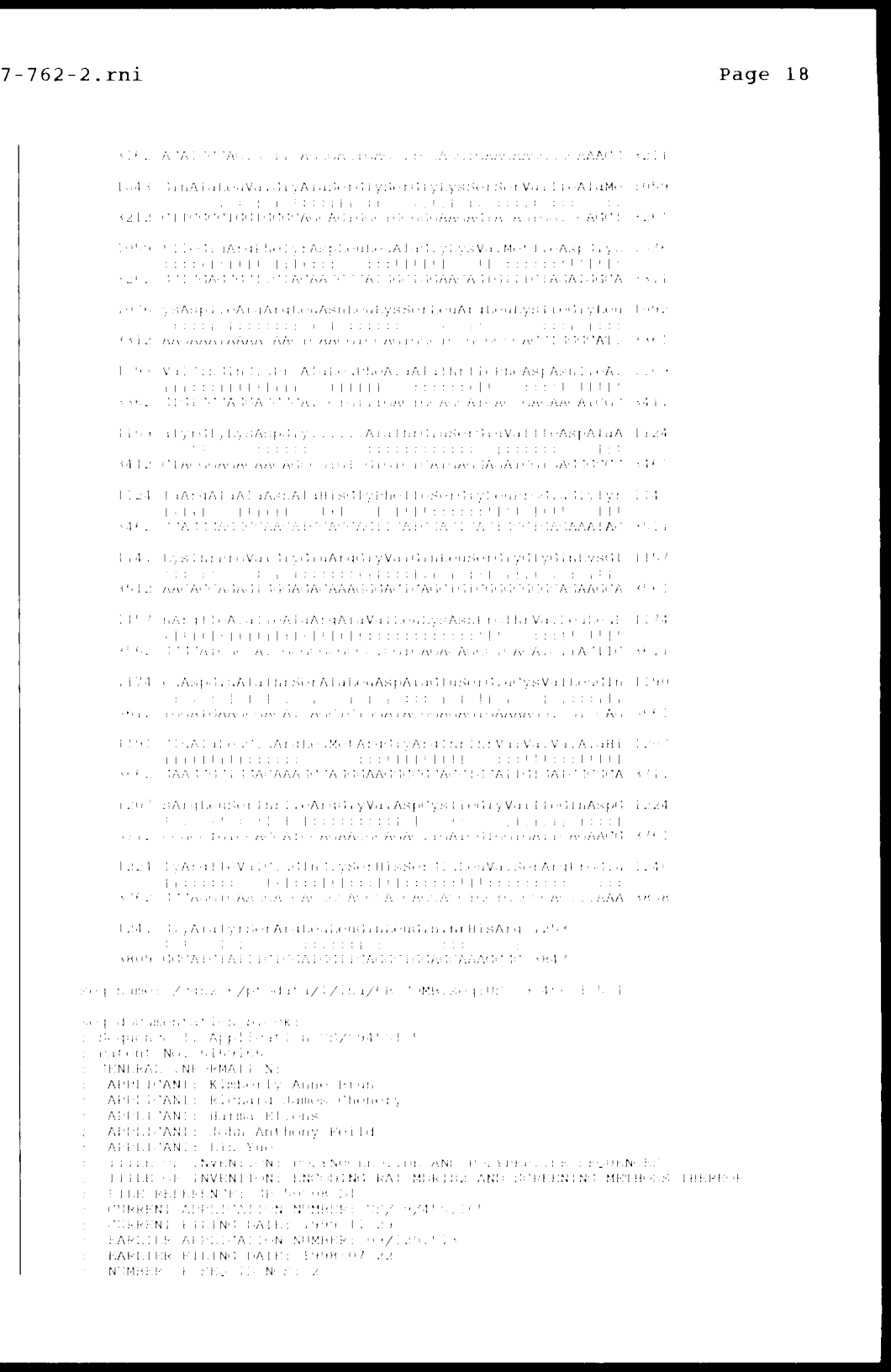
Alignment Block:
[illegible]

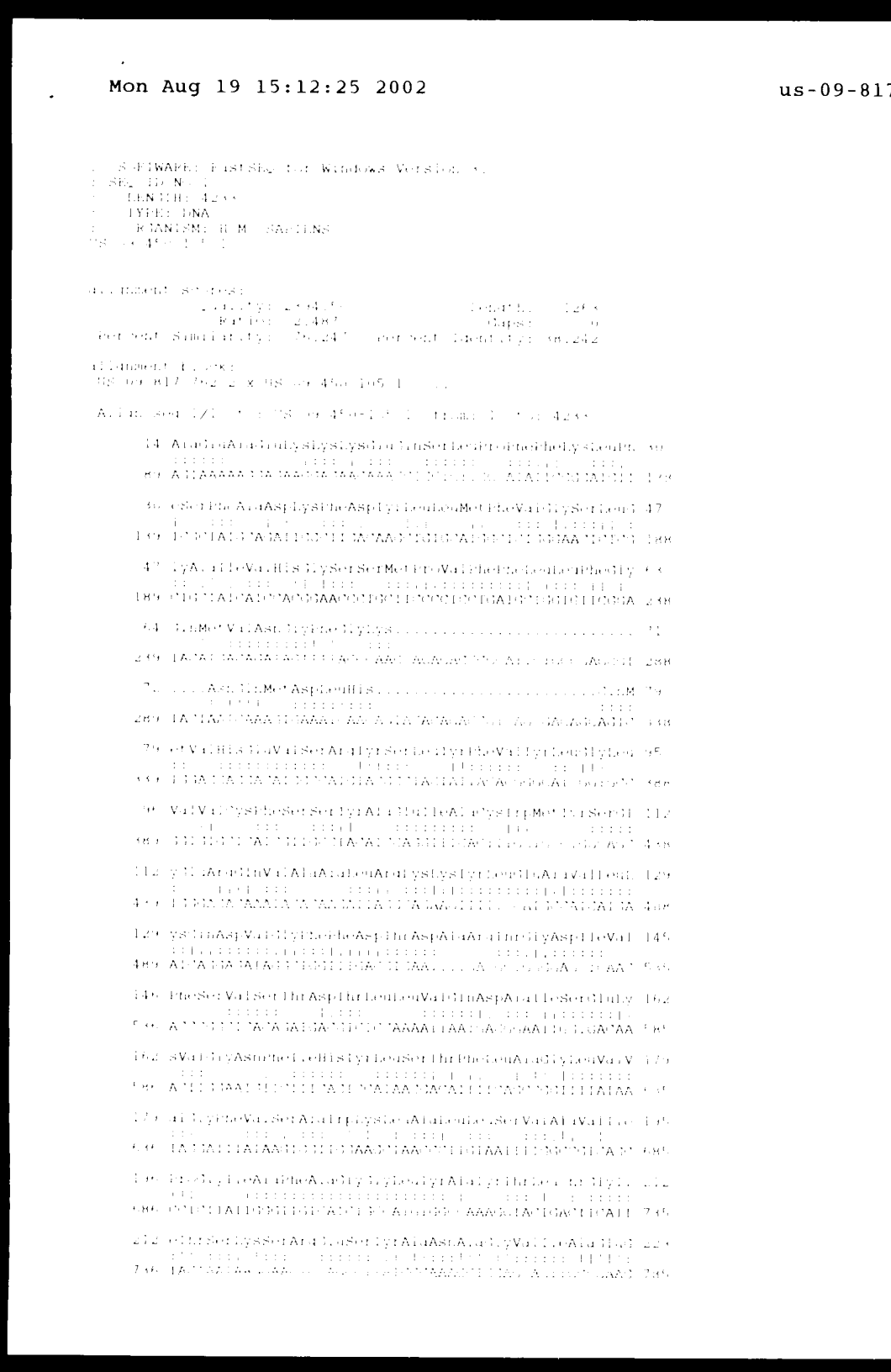
[illegible]

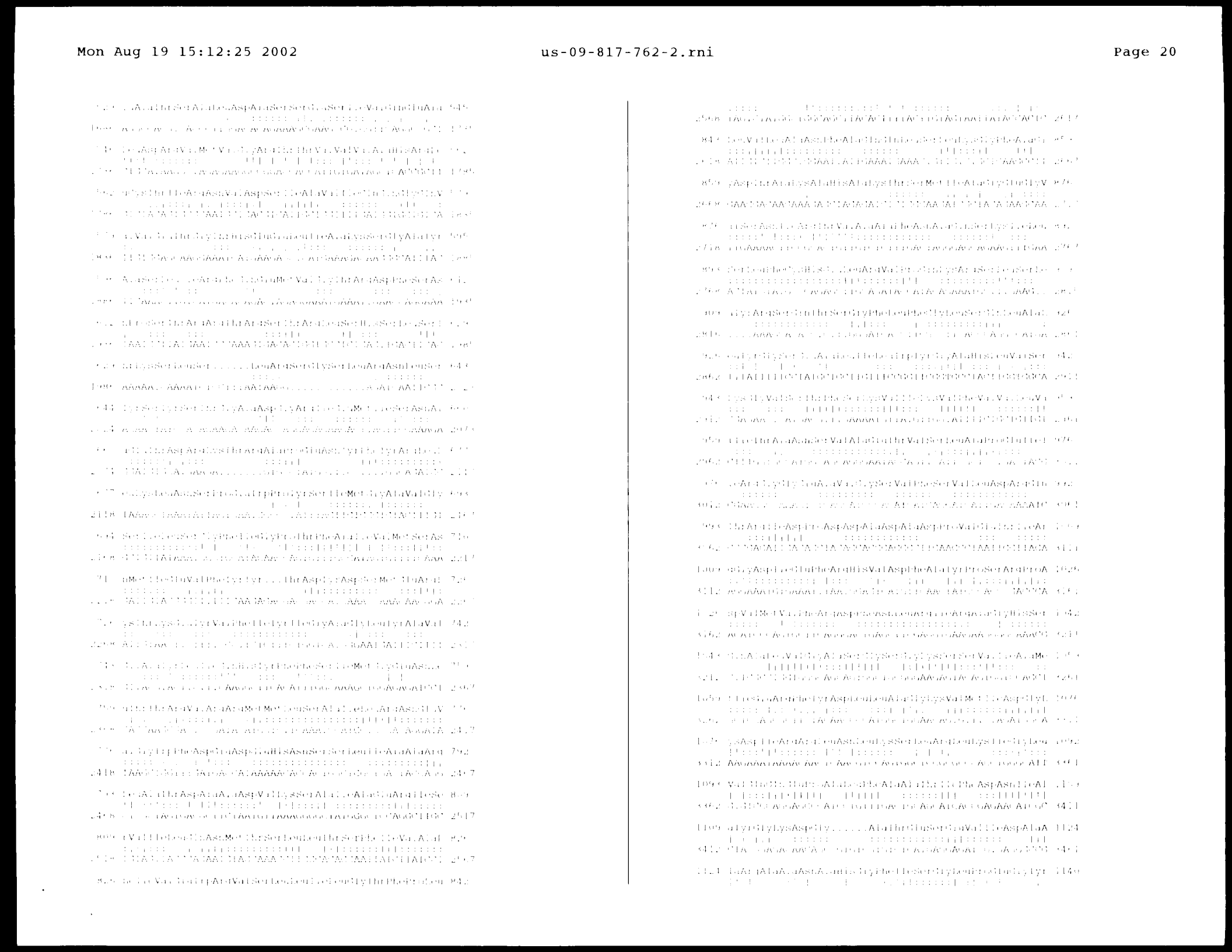


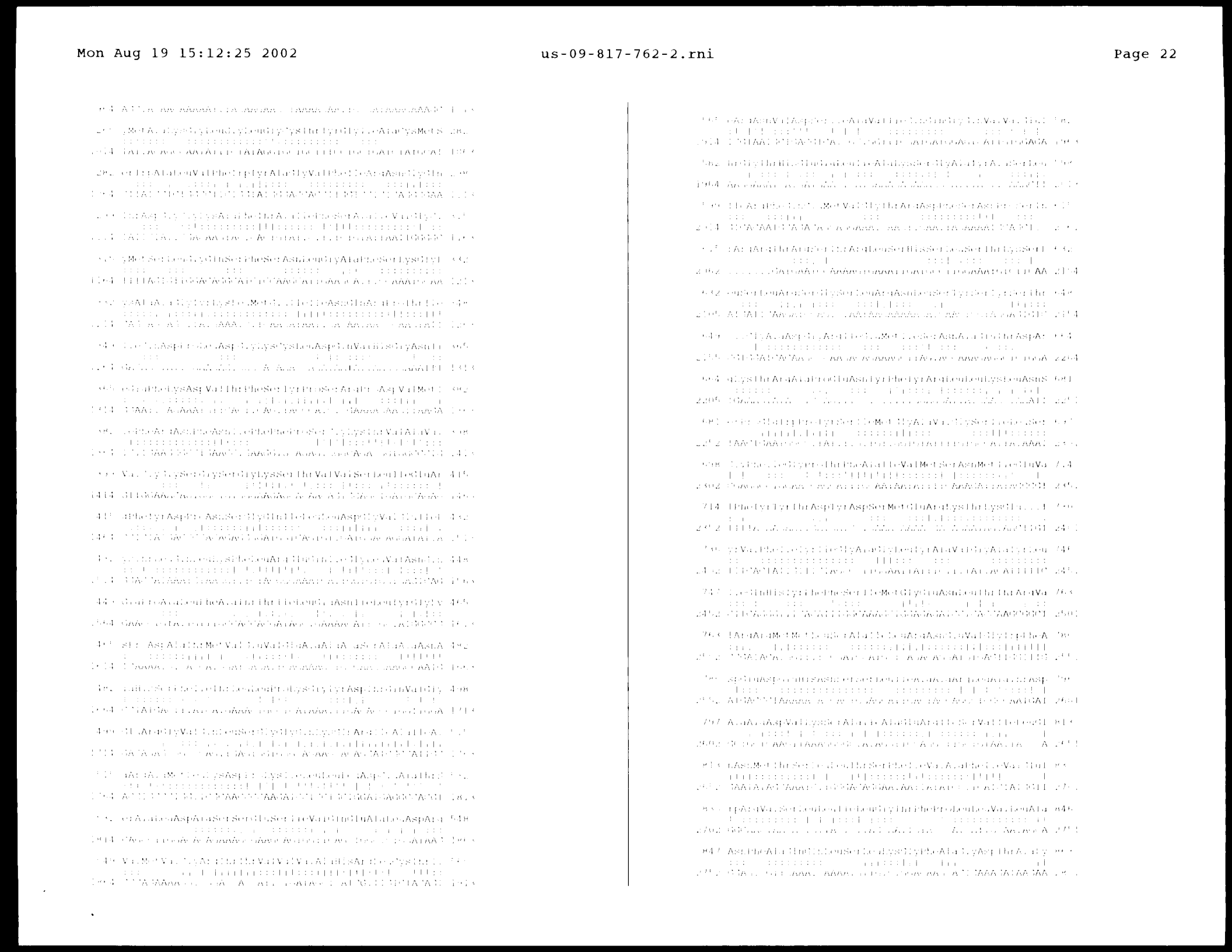


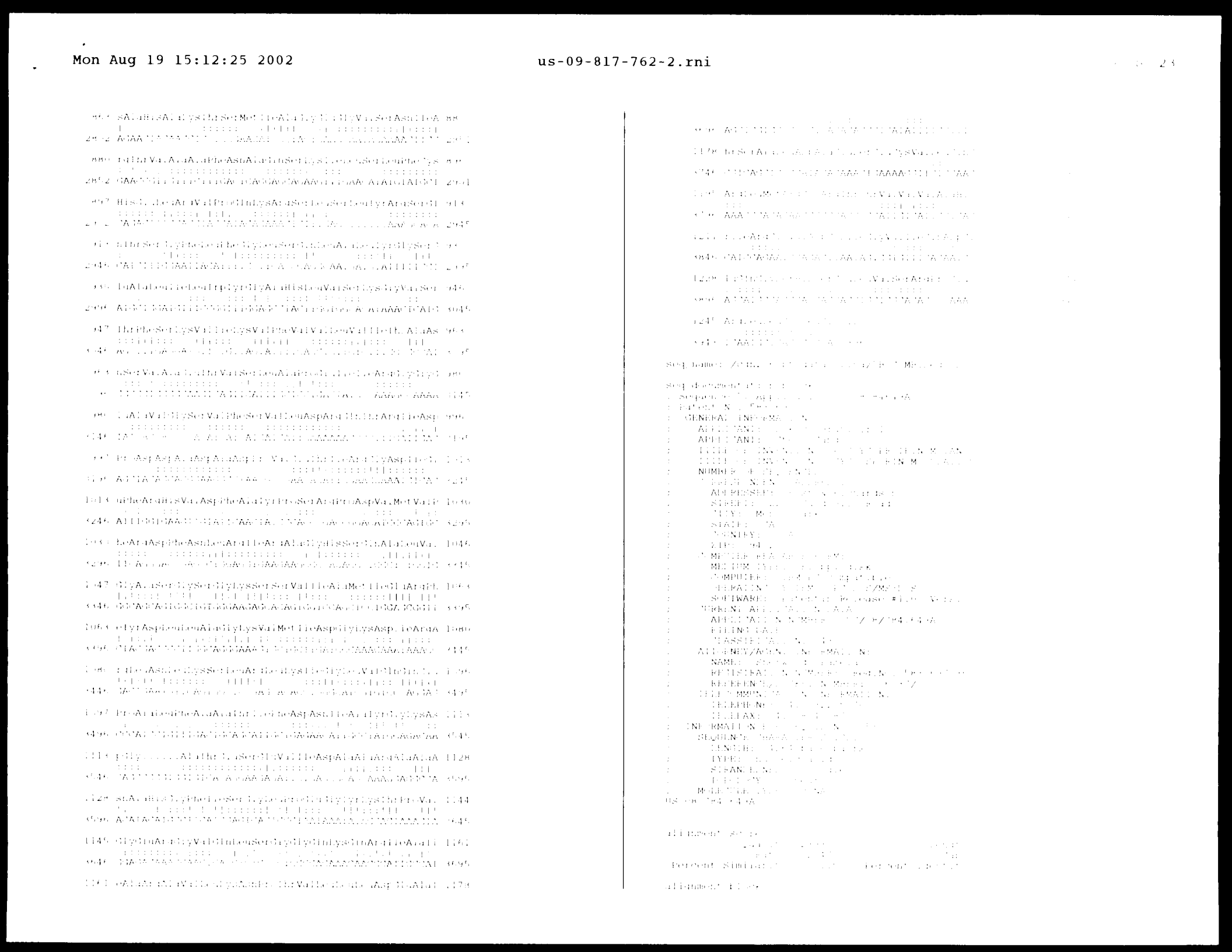


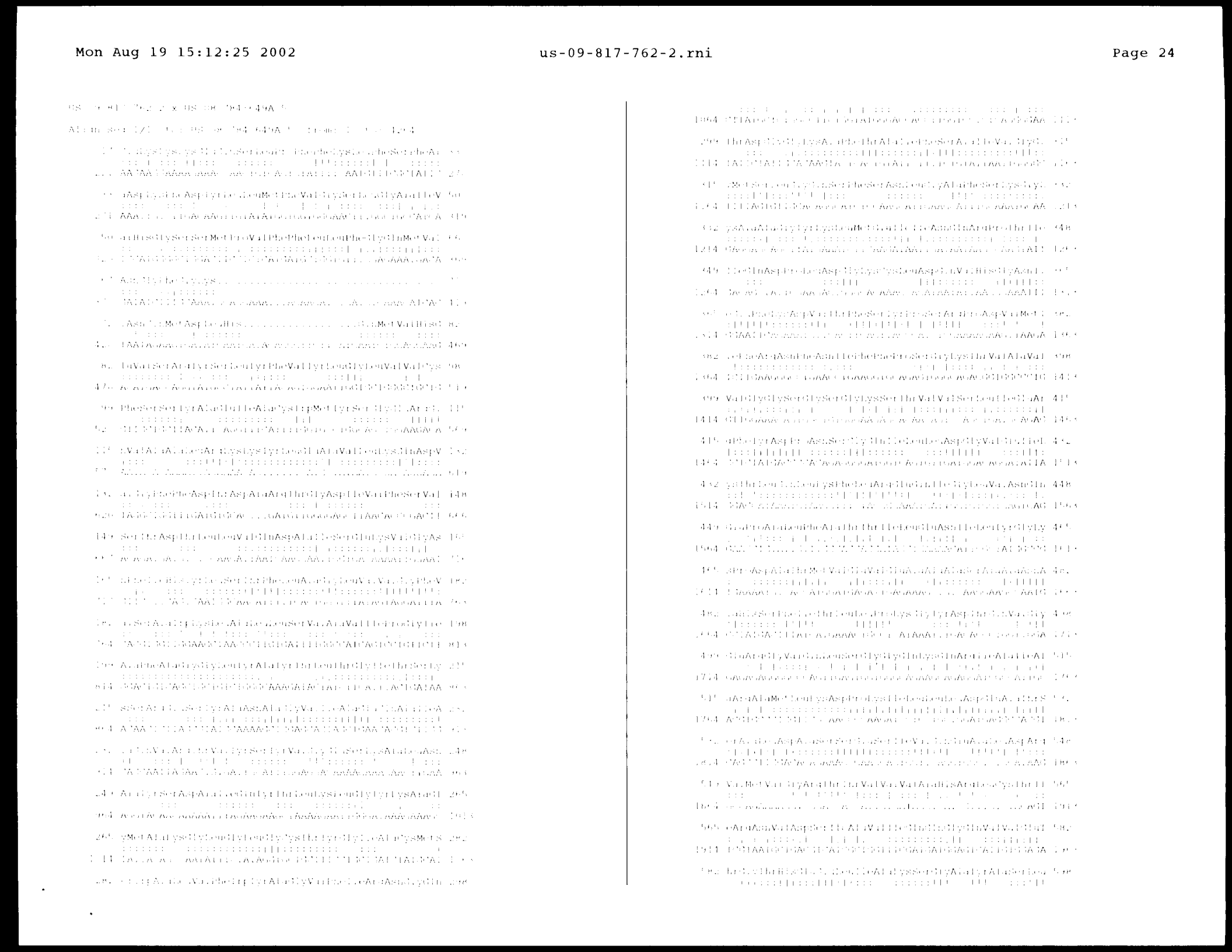


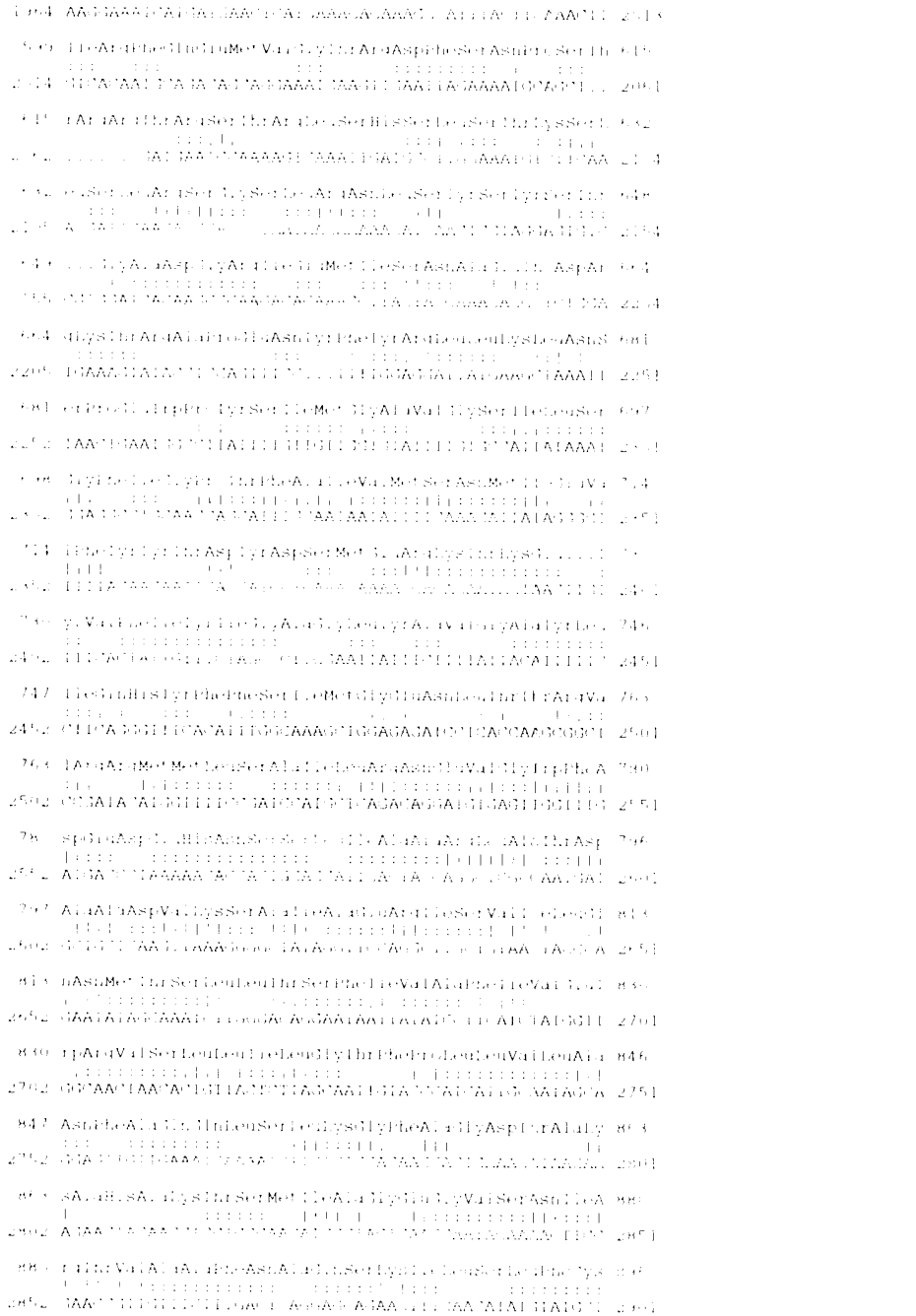


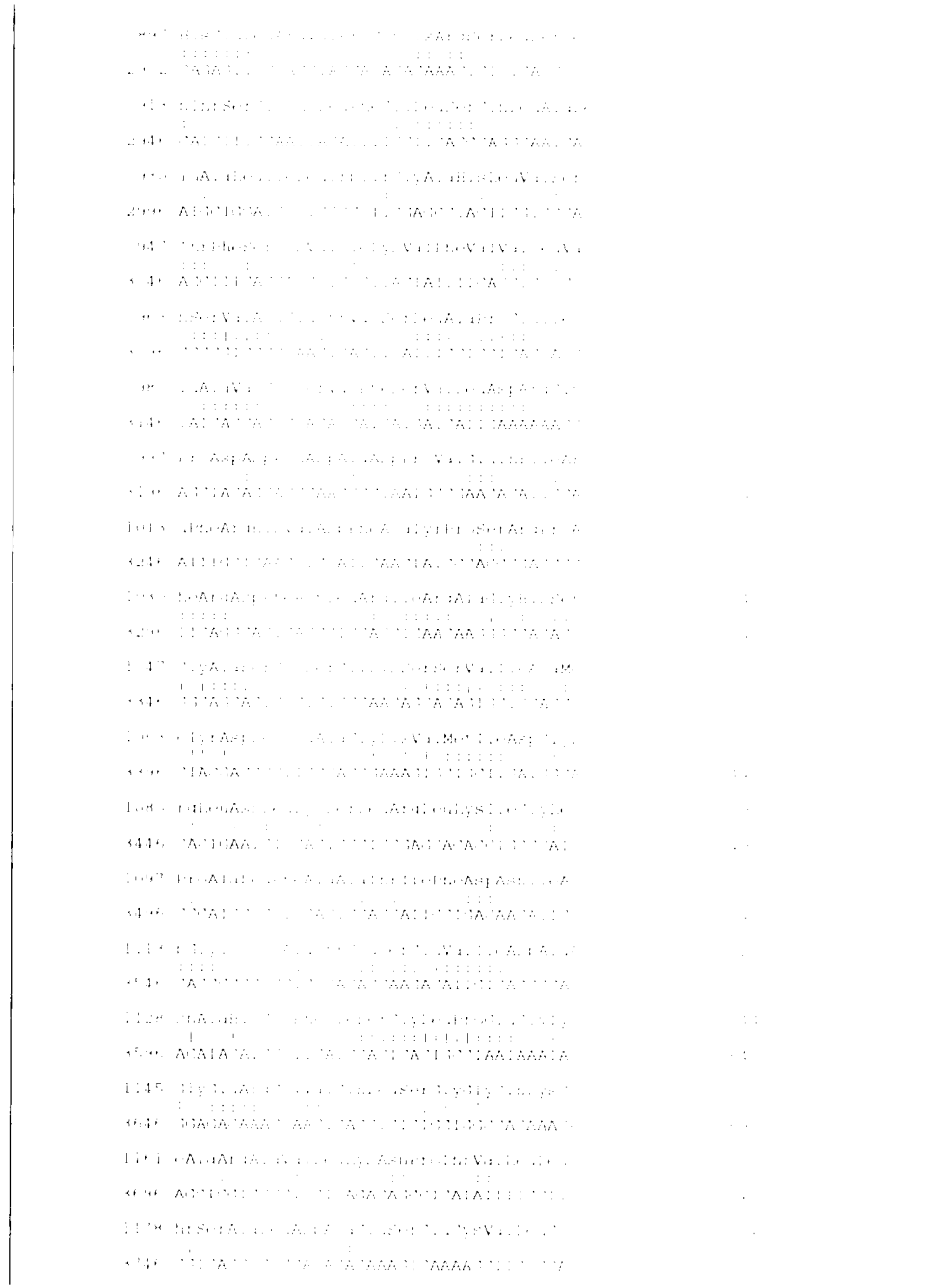


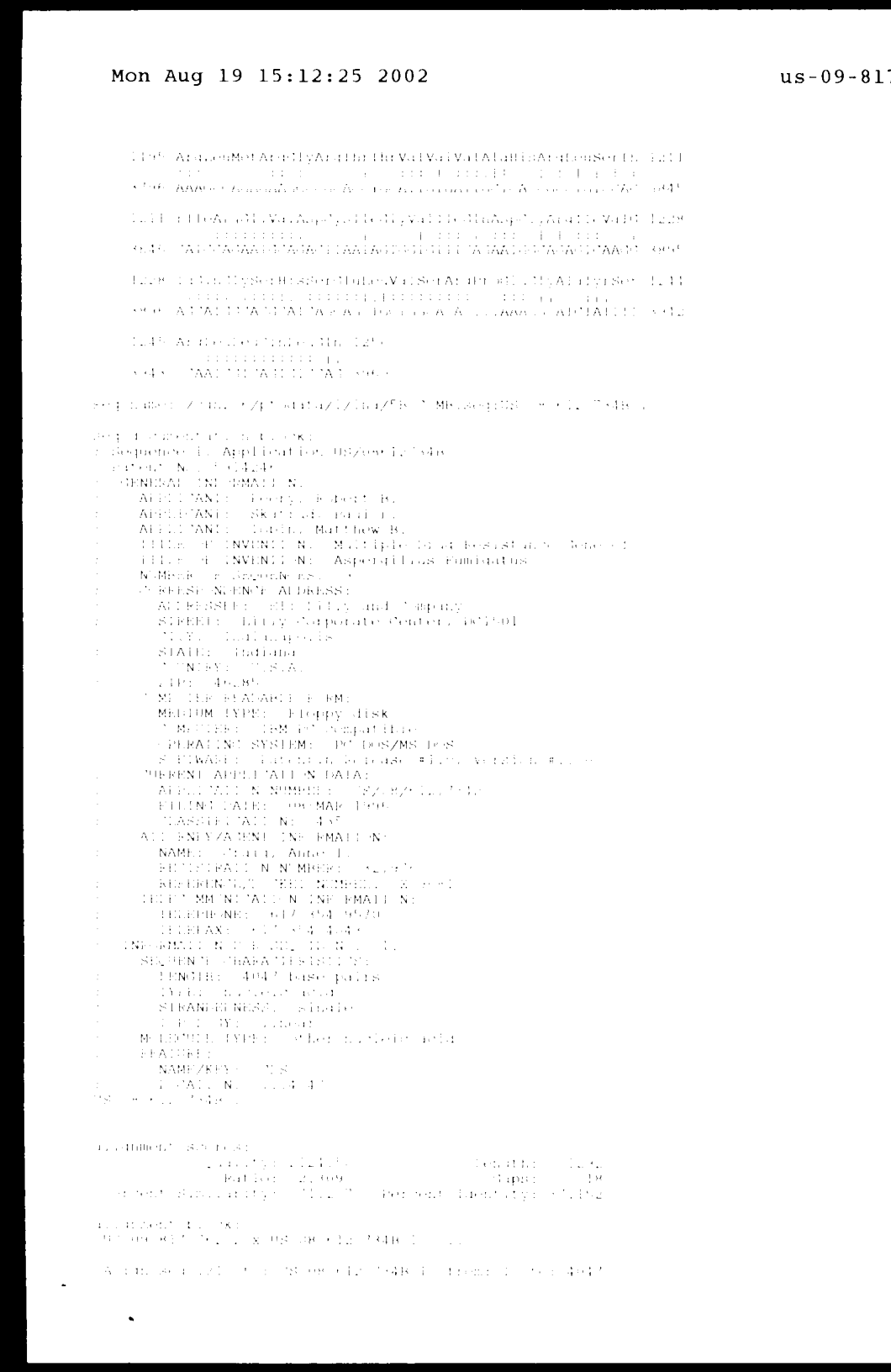


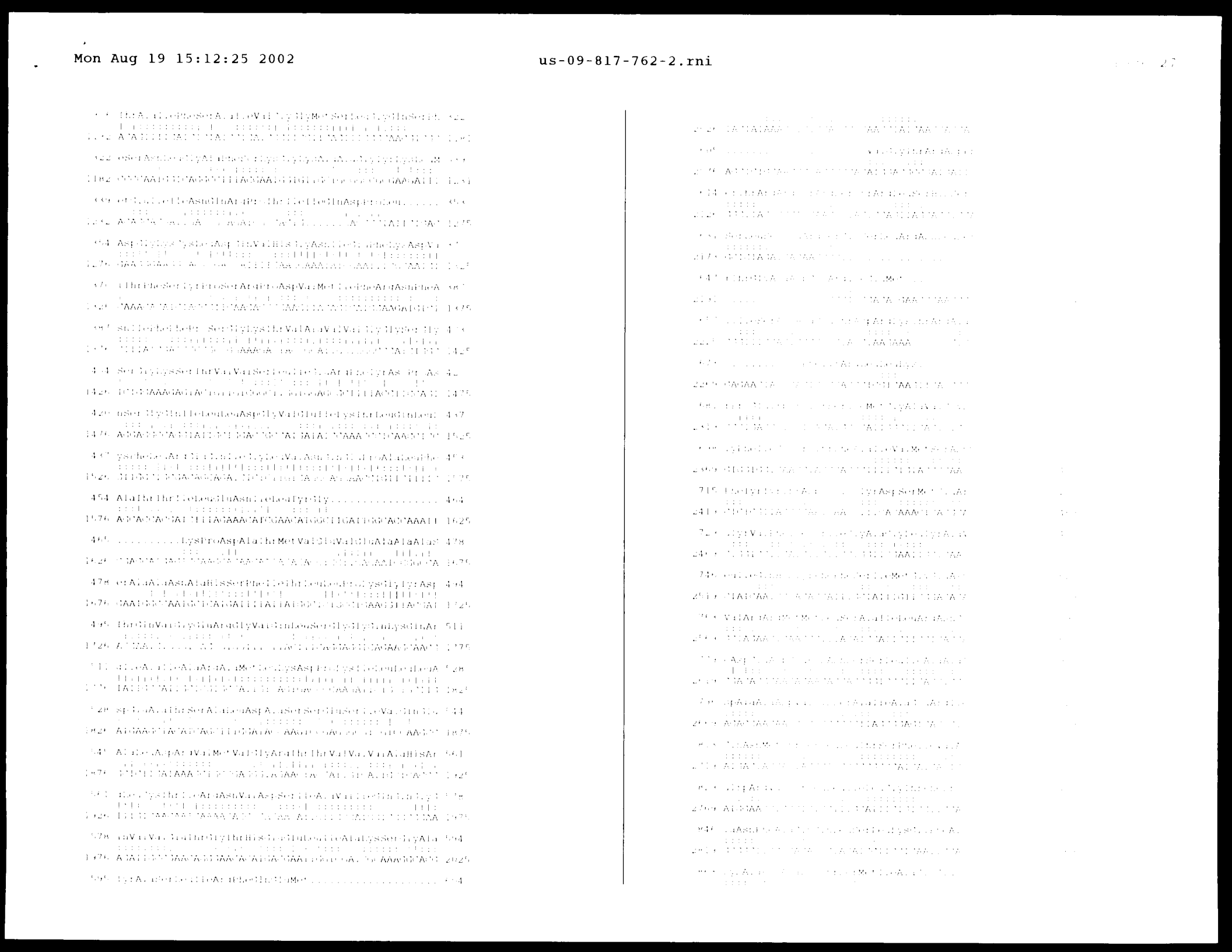


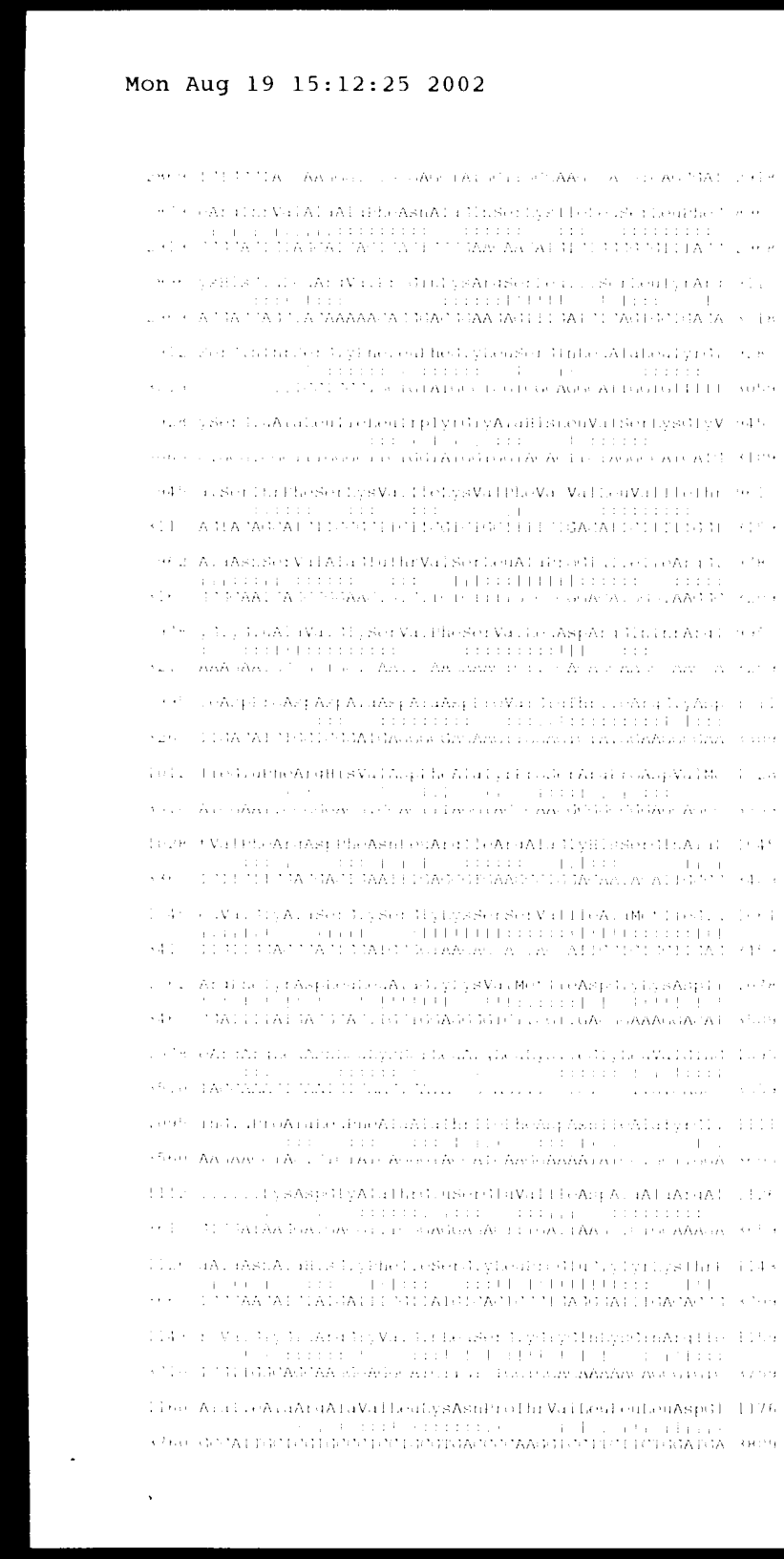


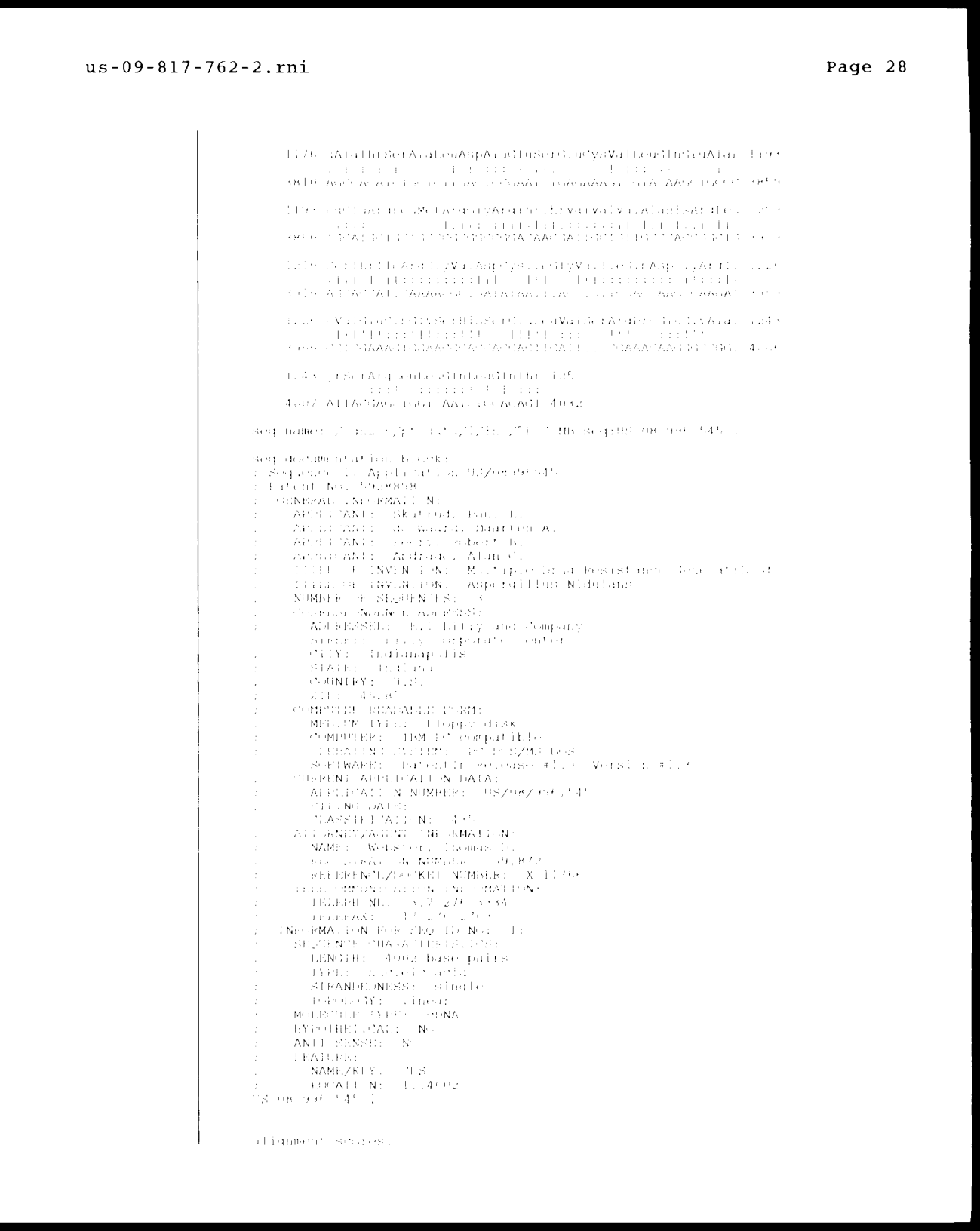


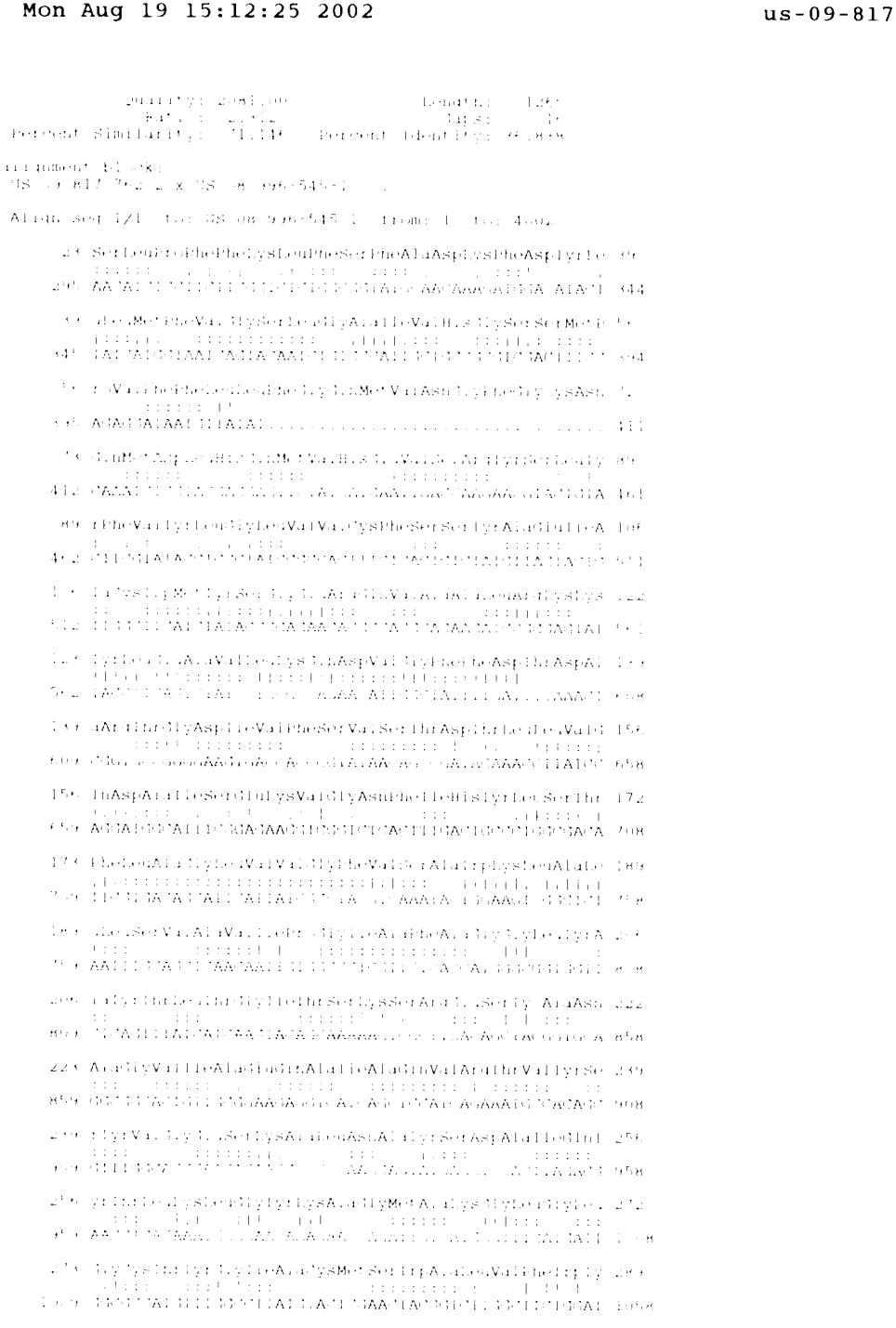


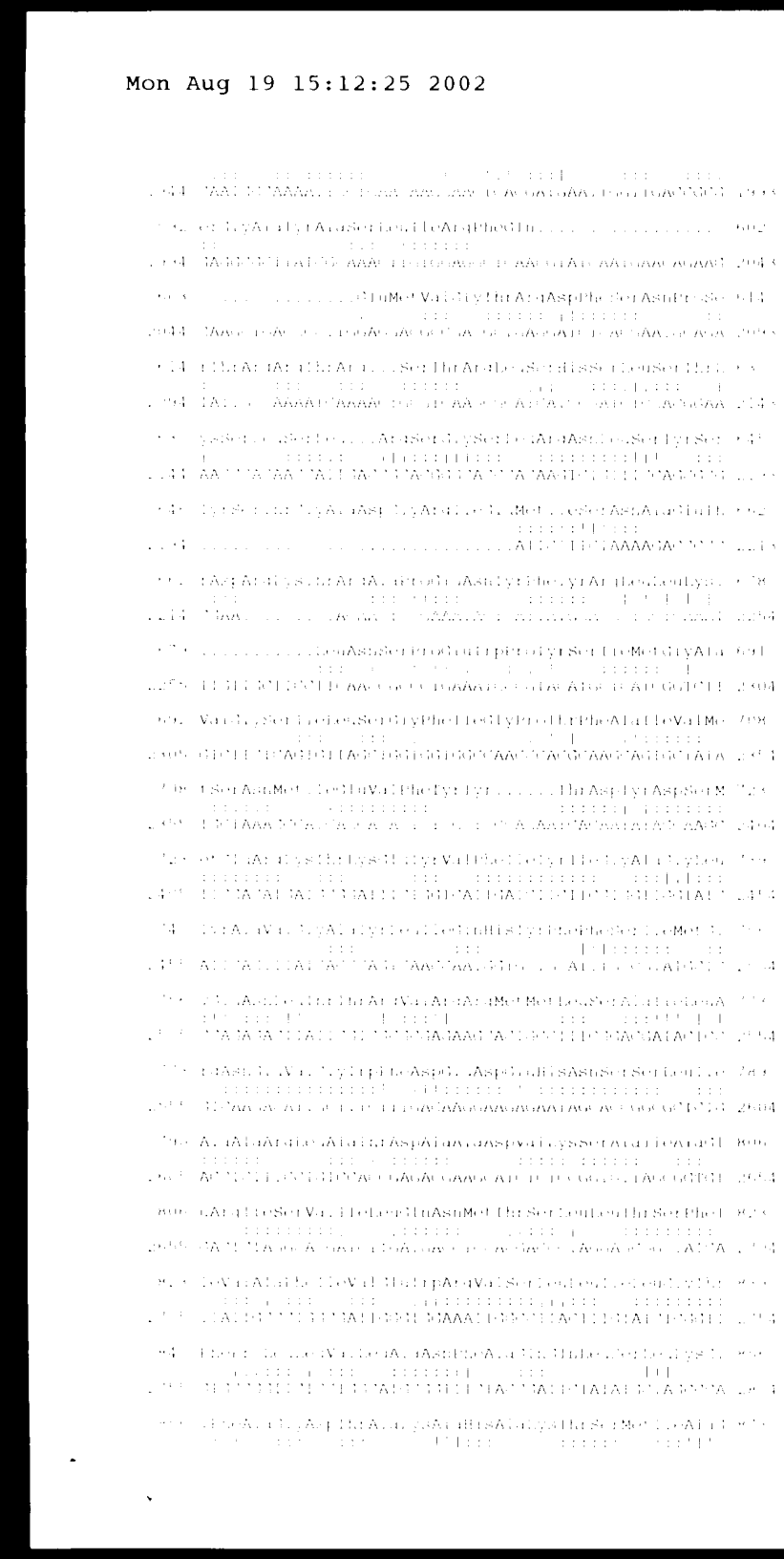


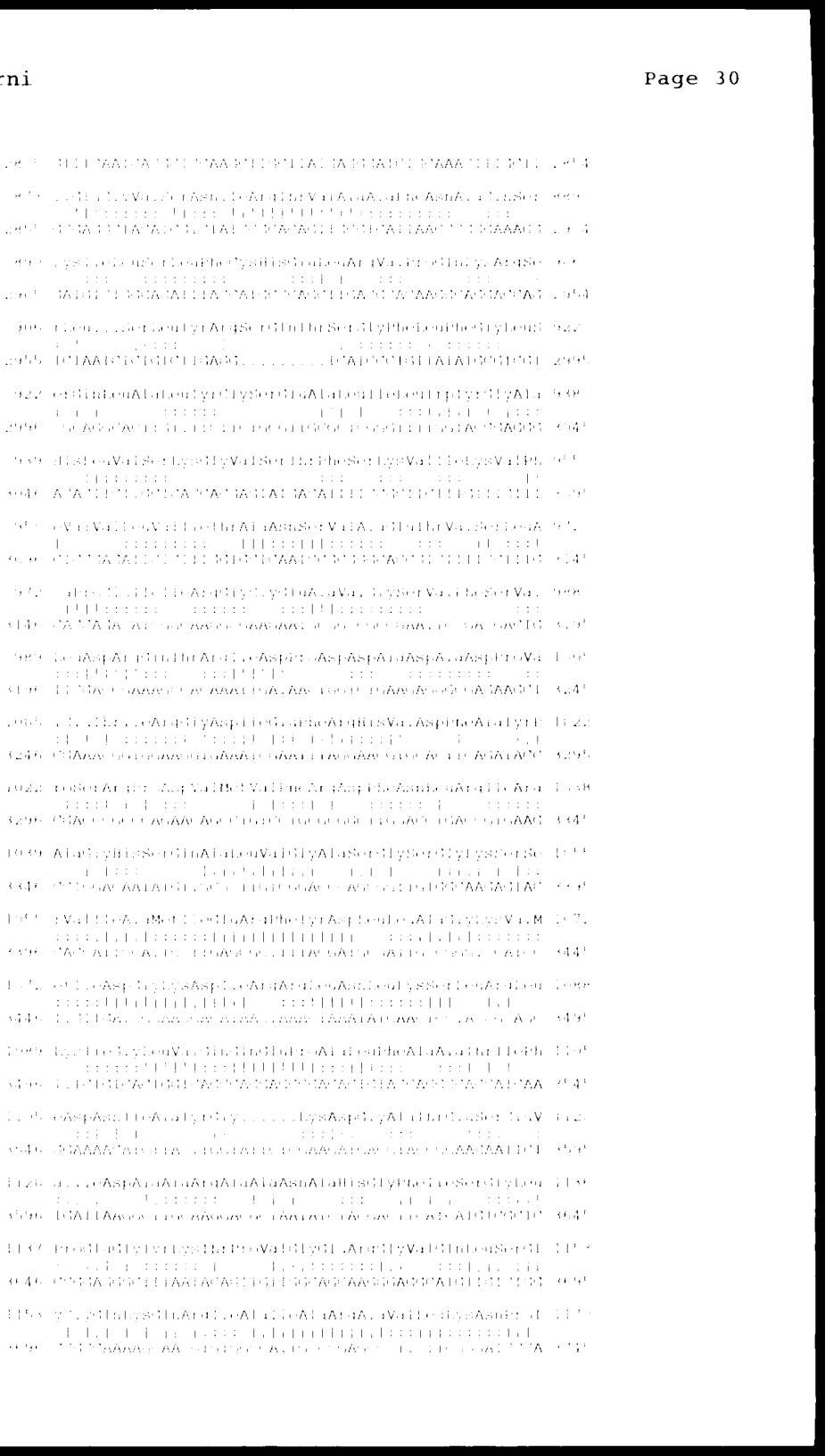


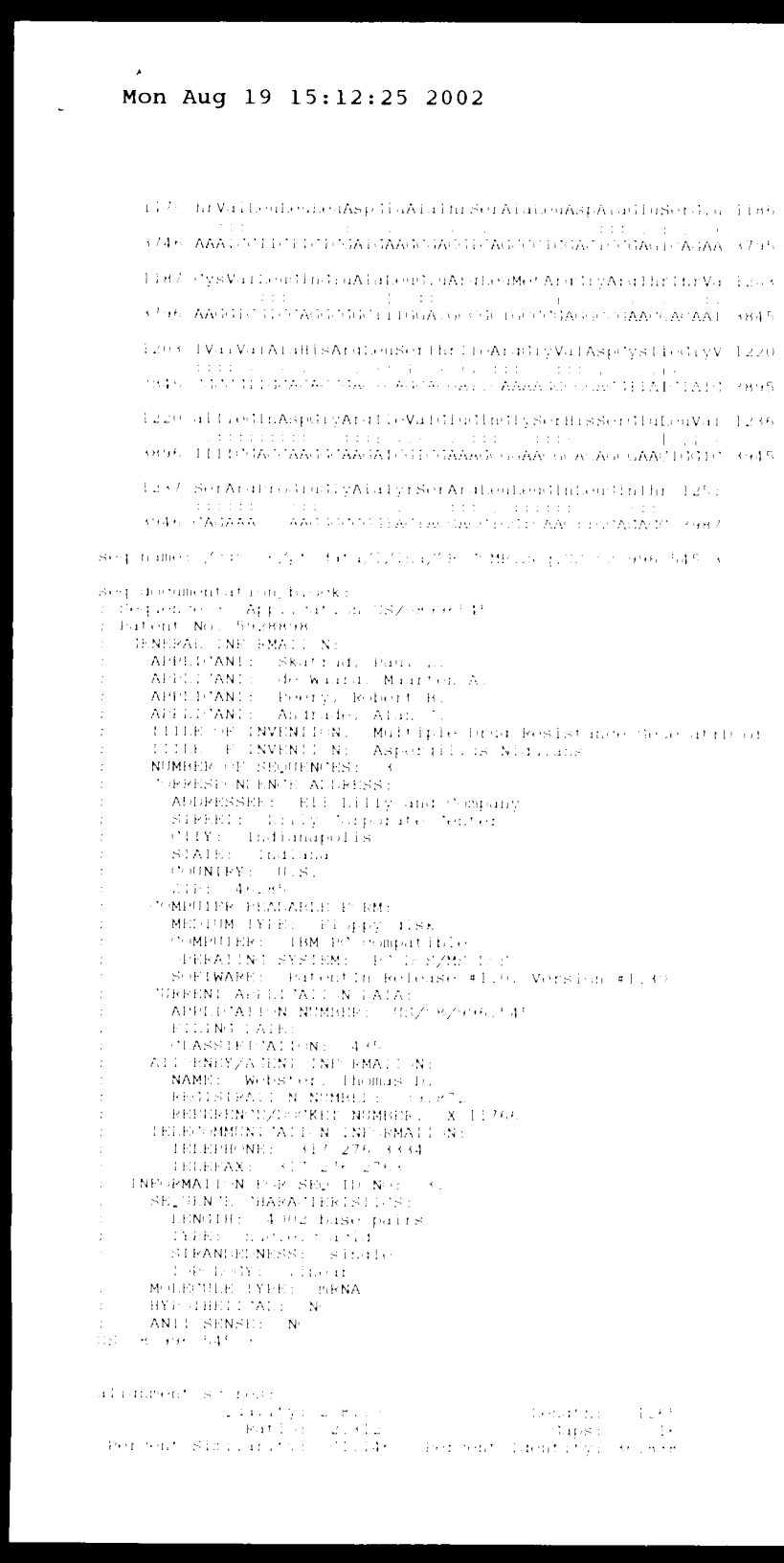


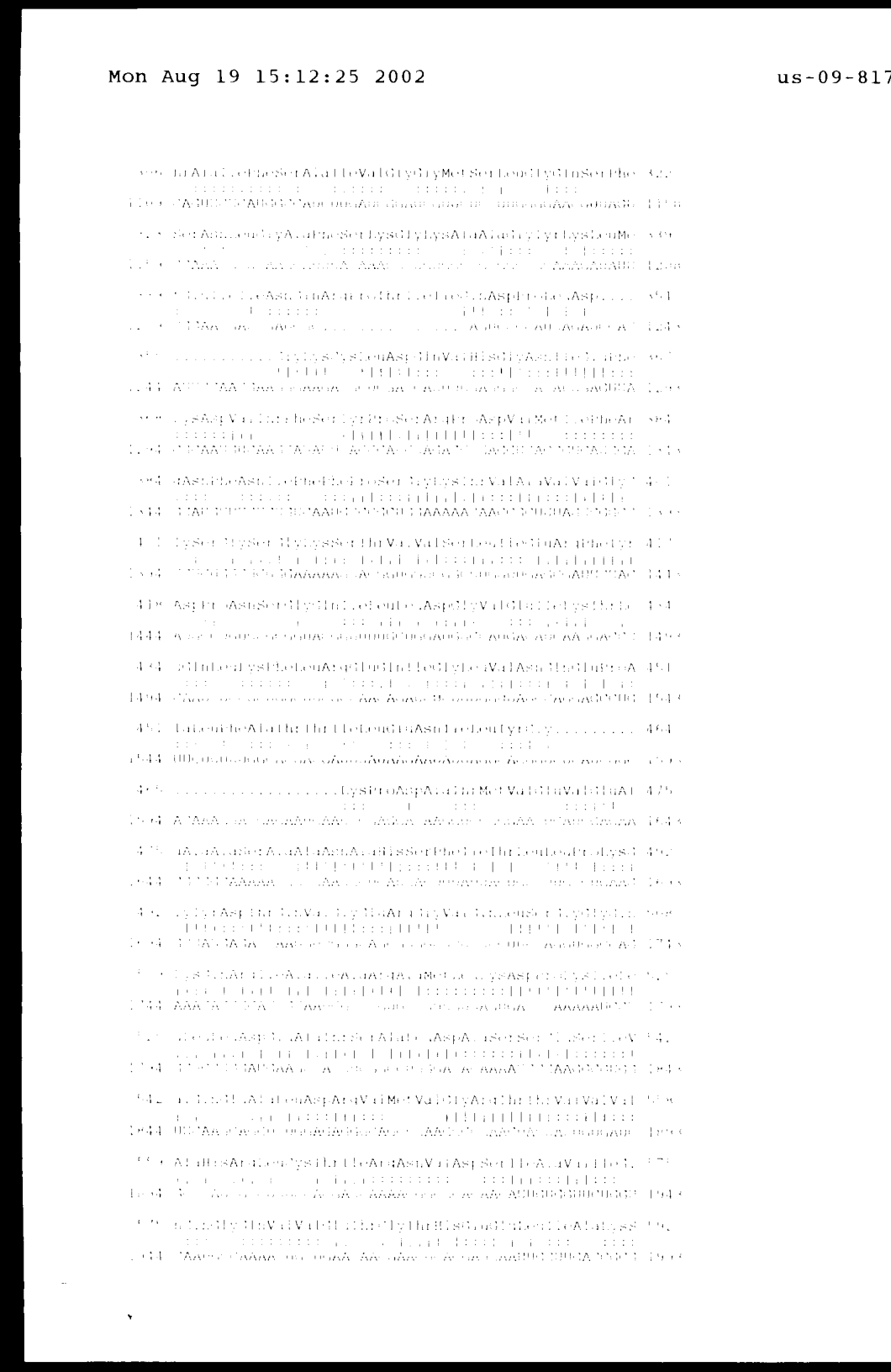


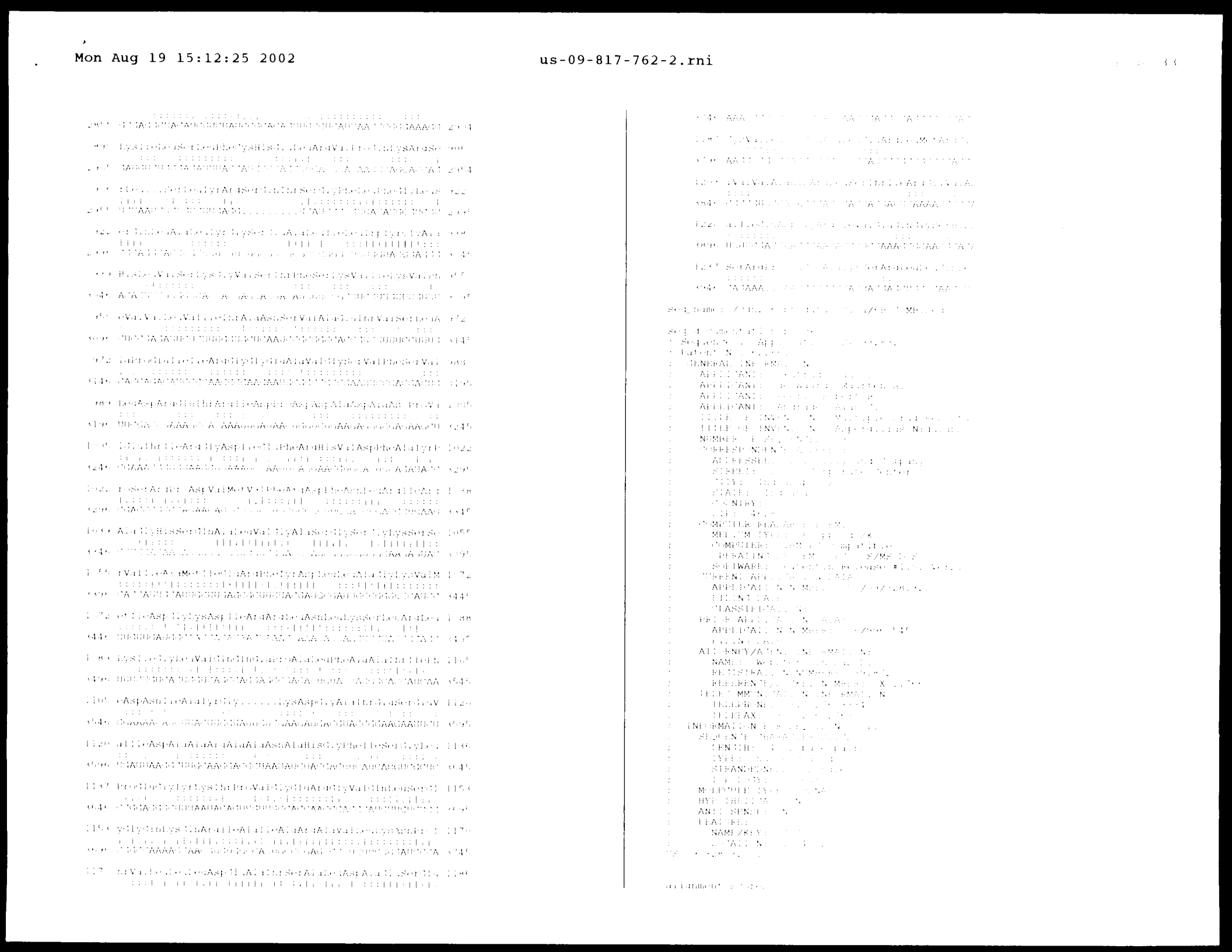


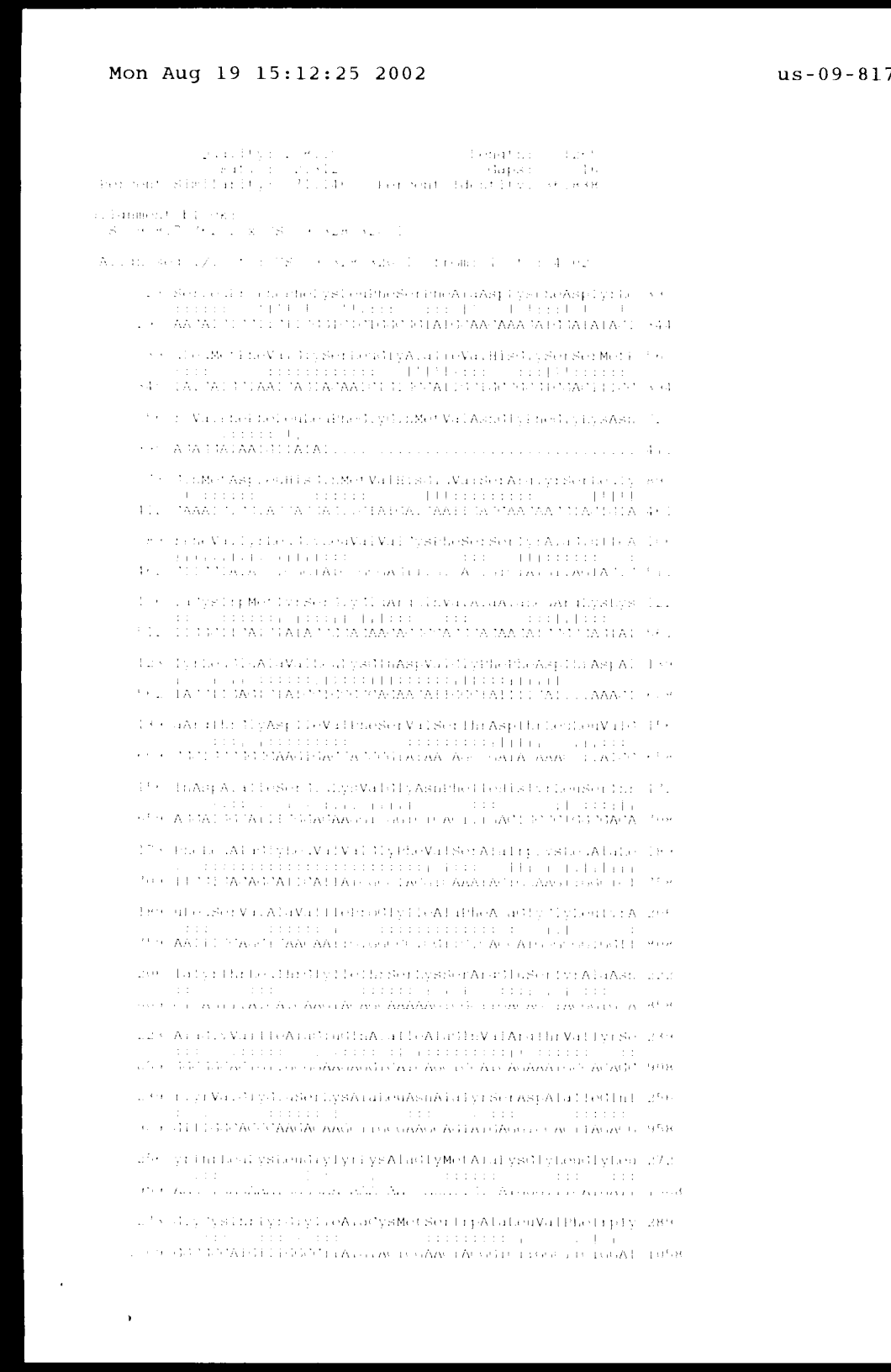


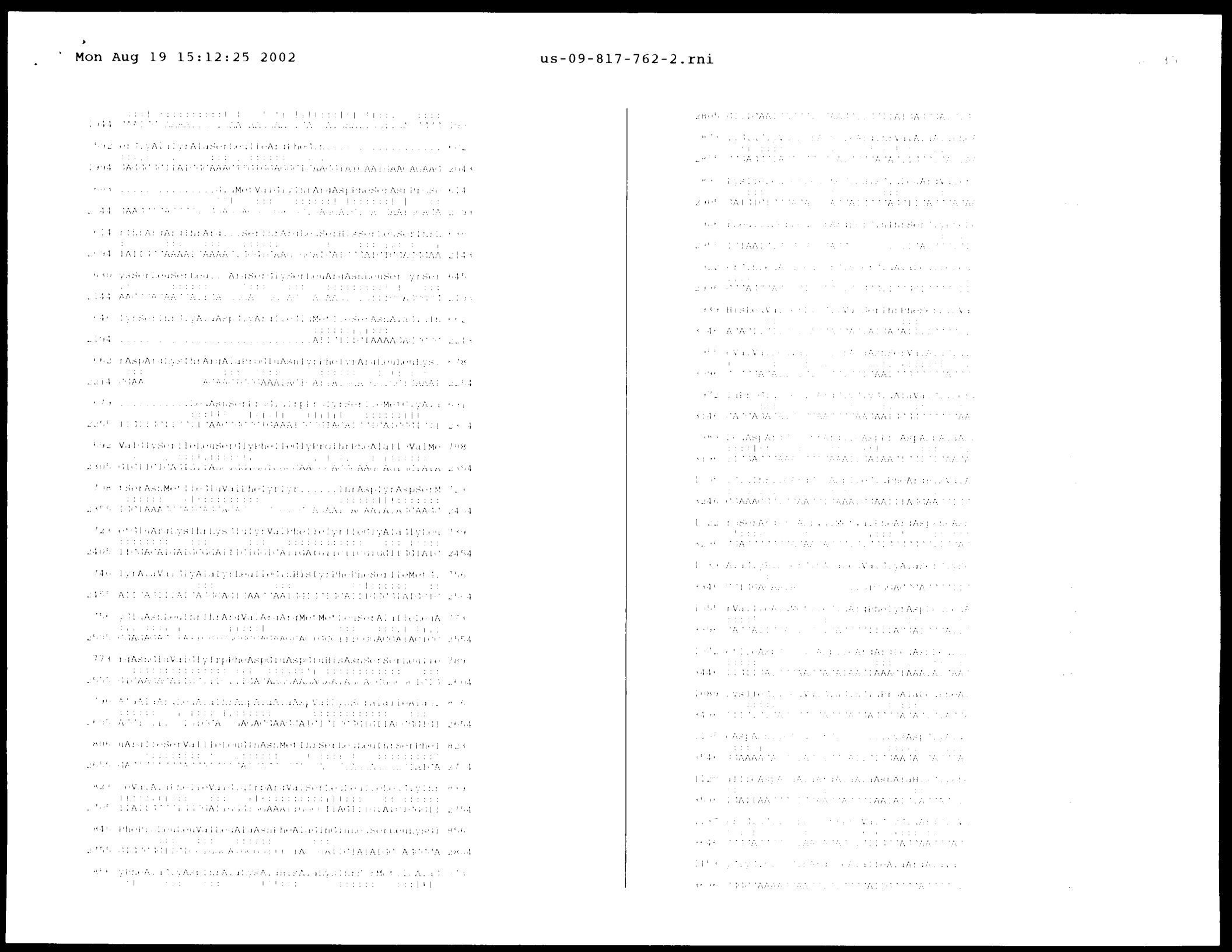


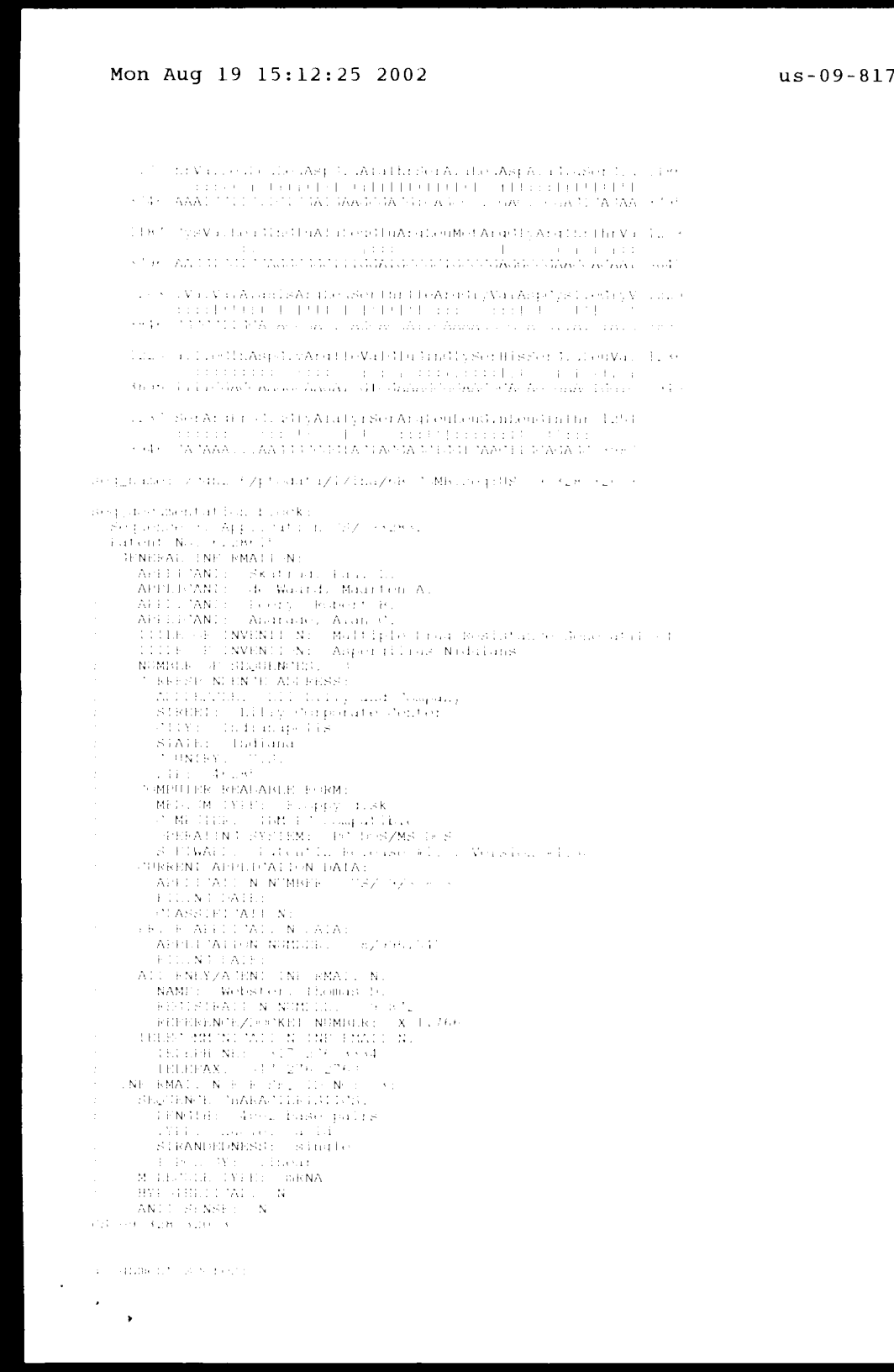


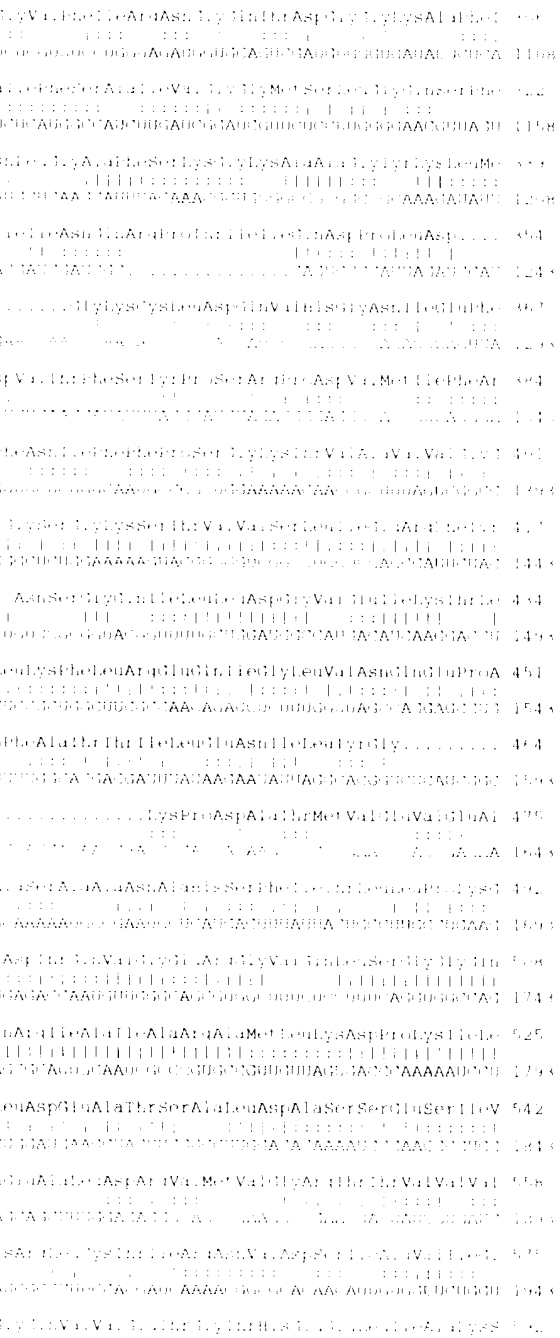


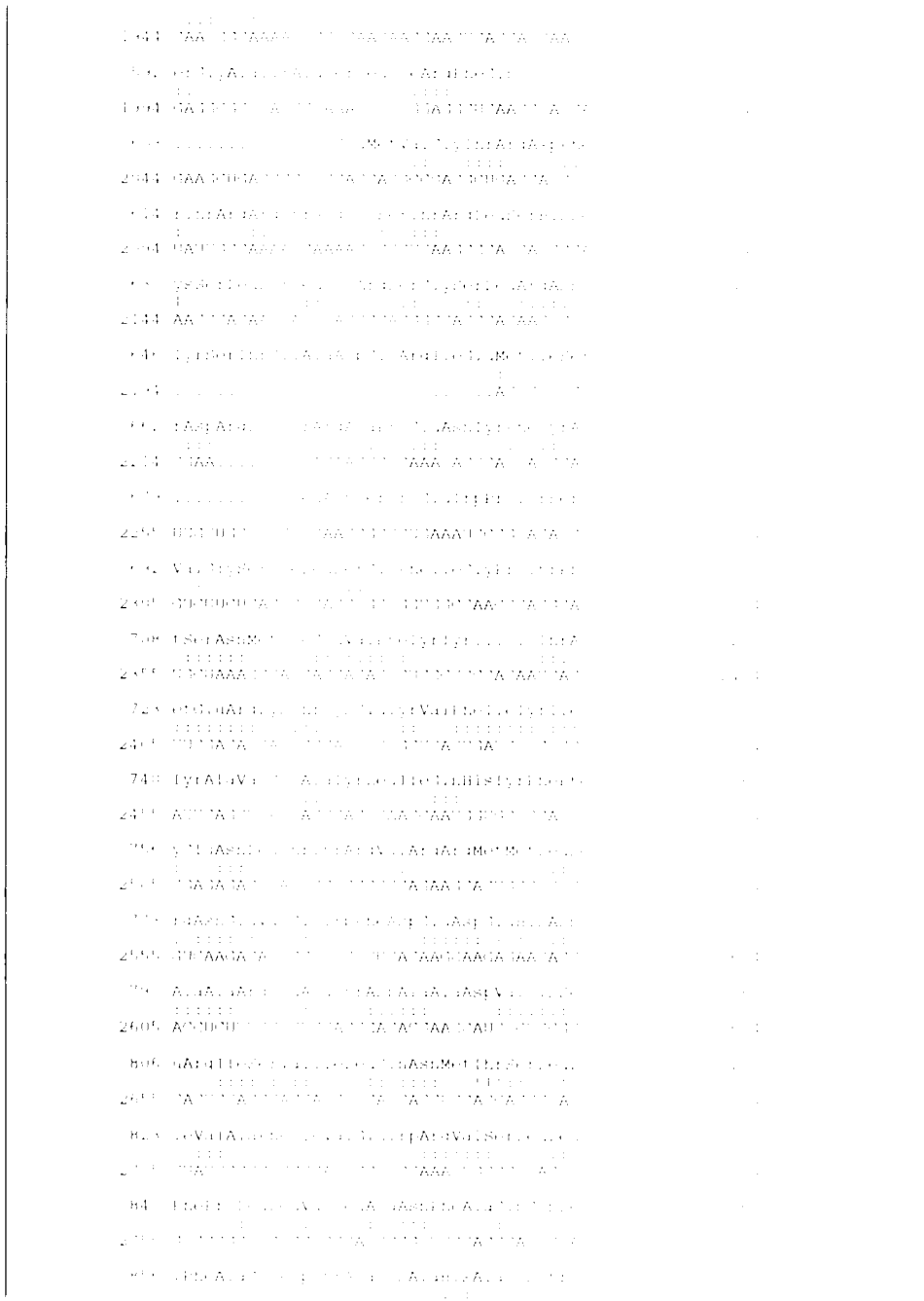


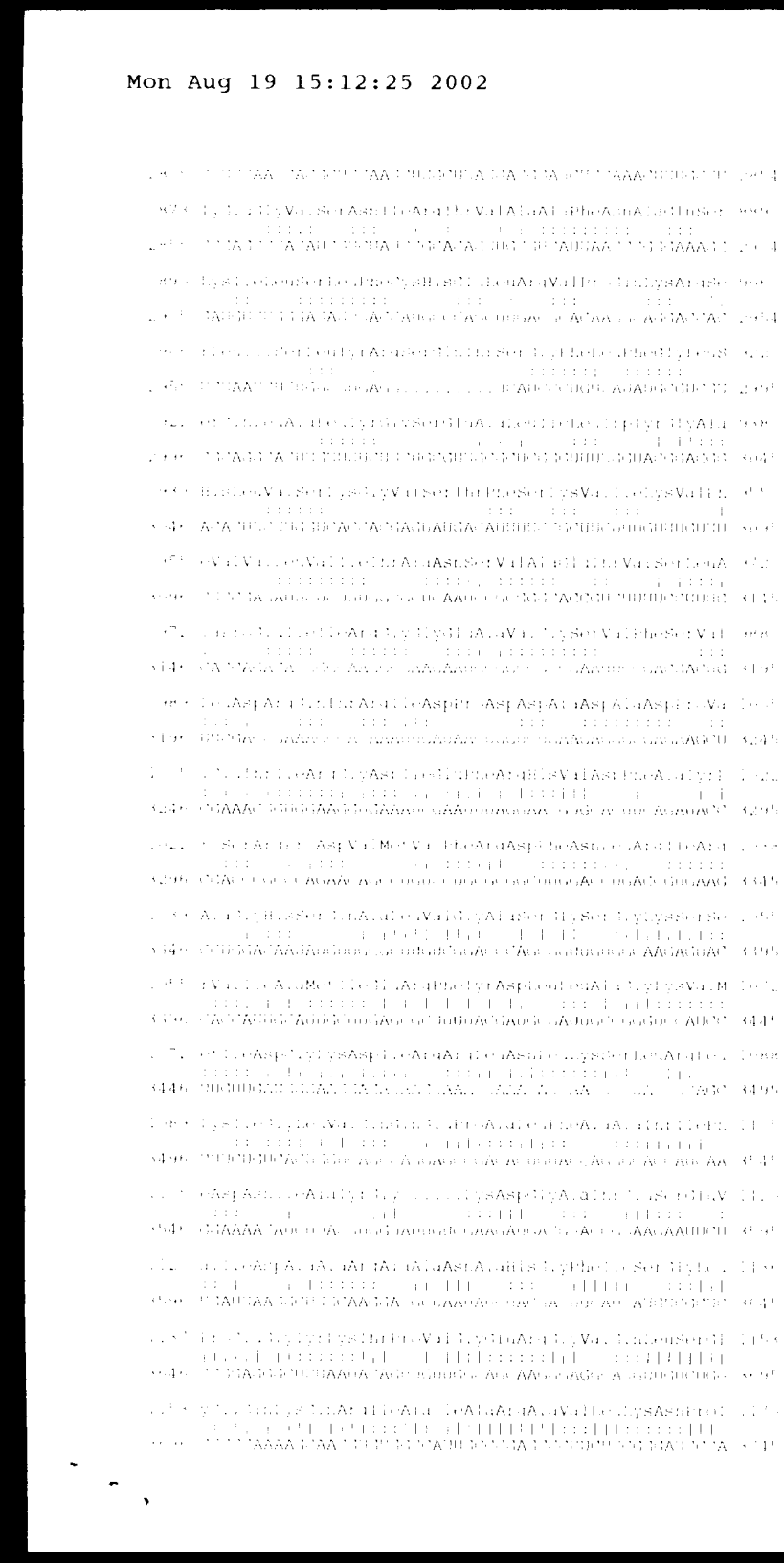


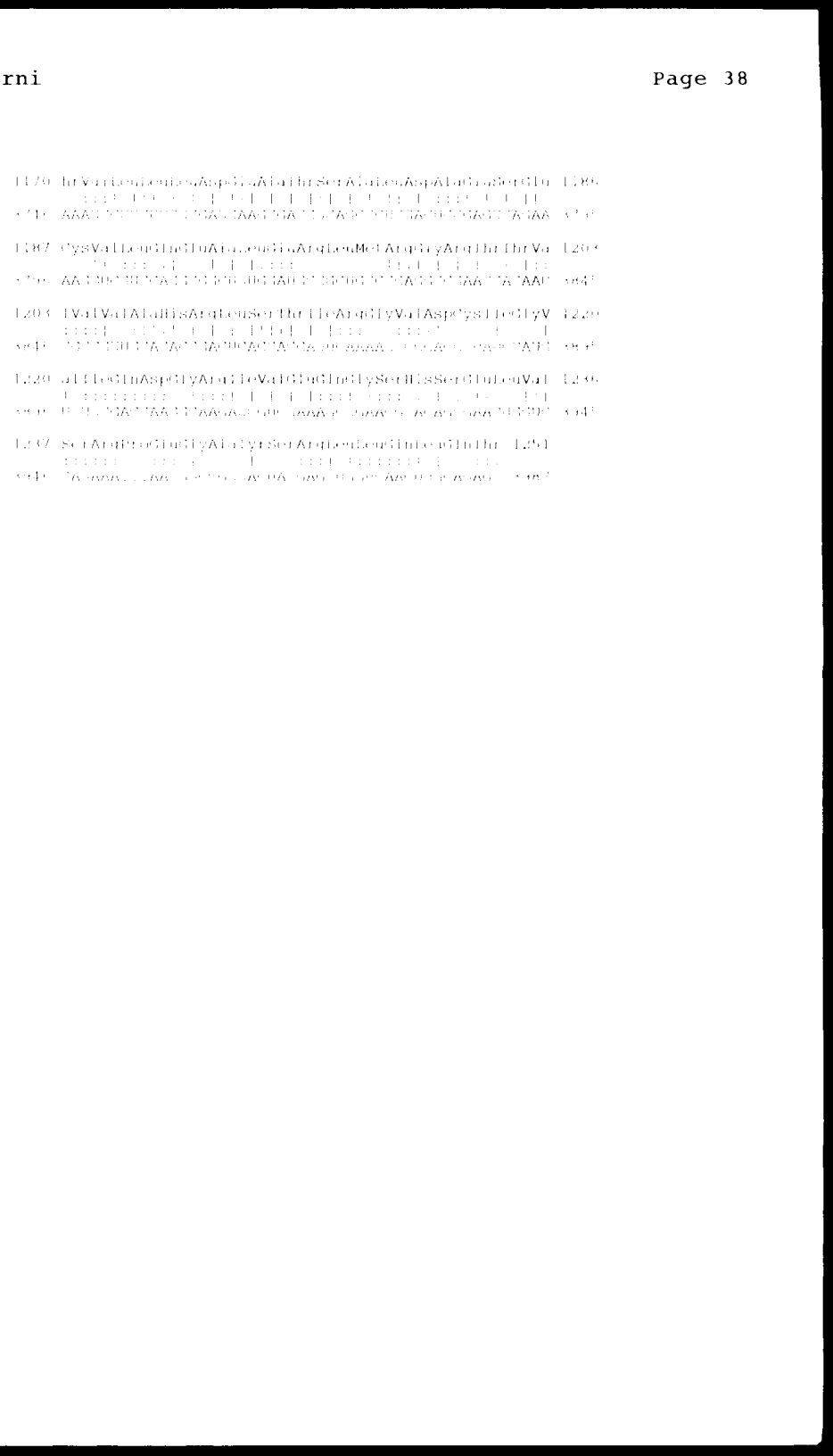


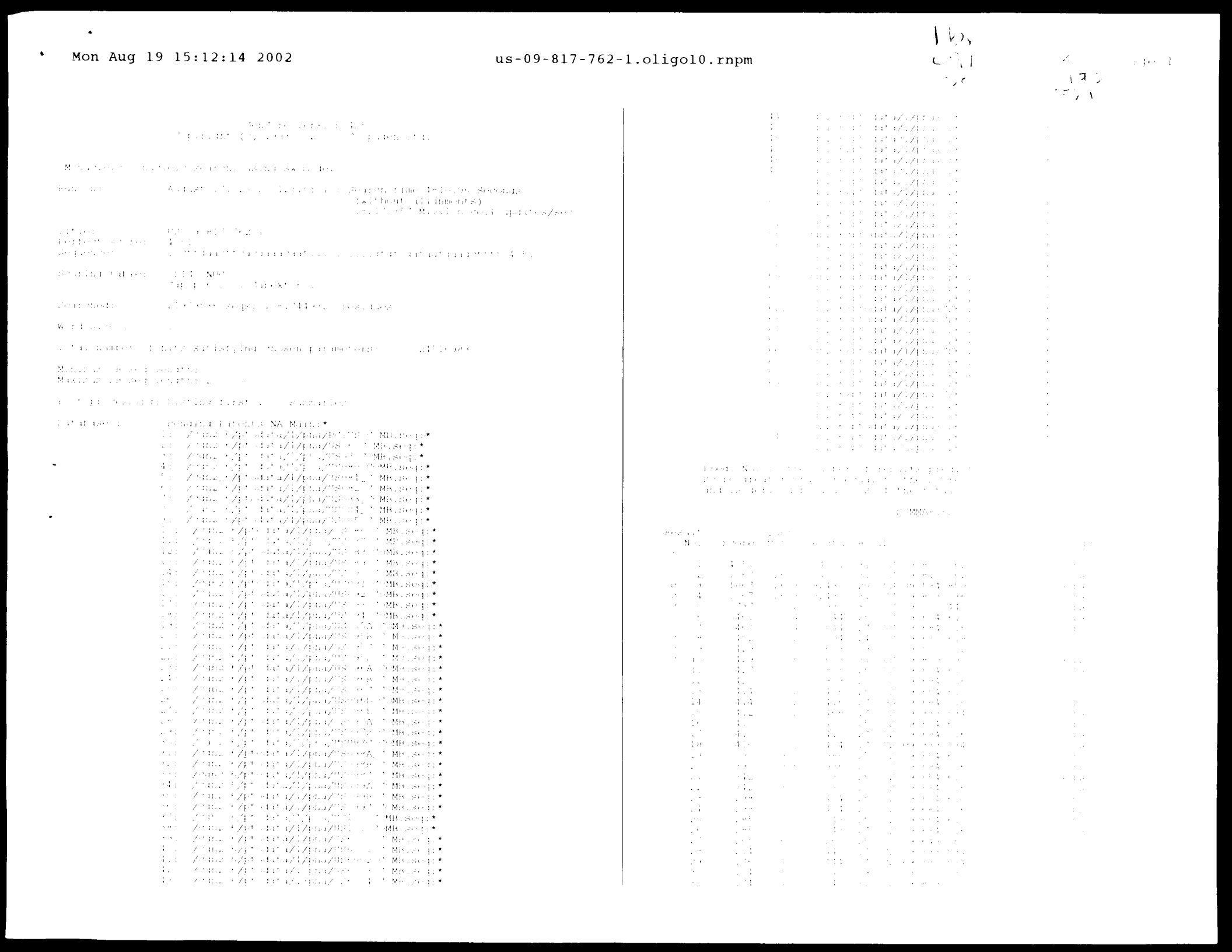


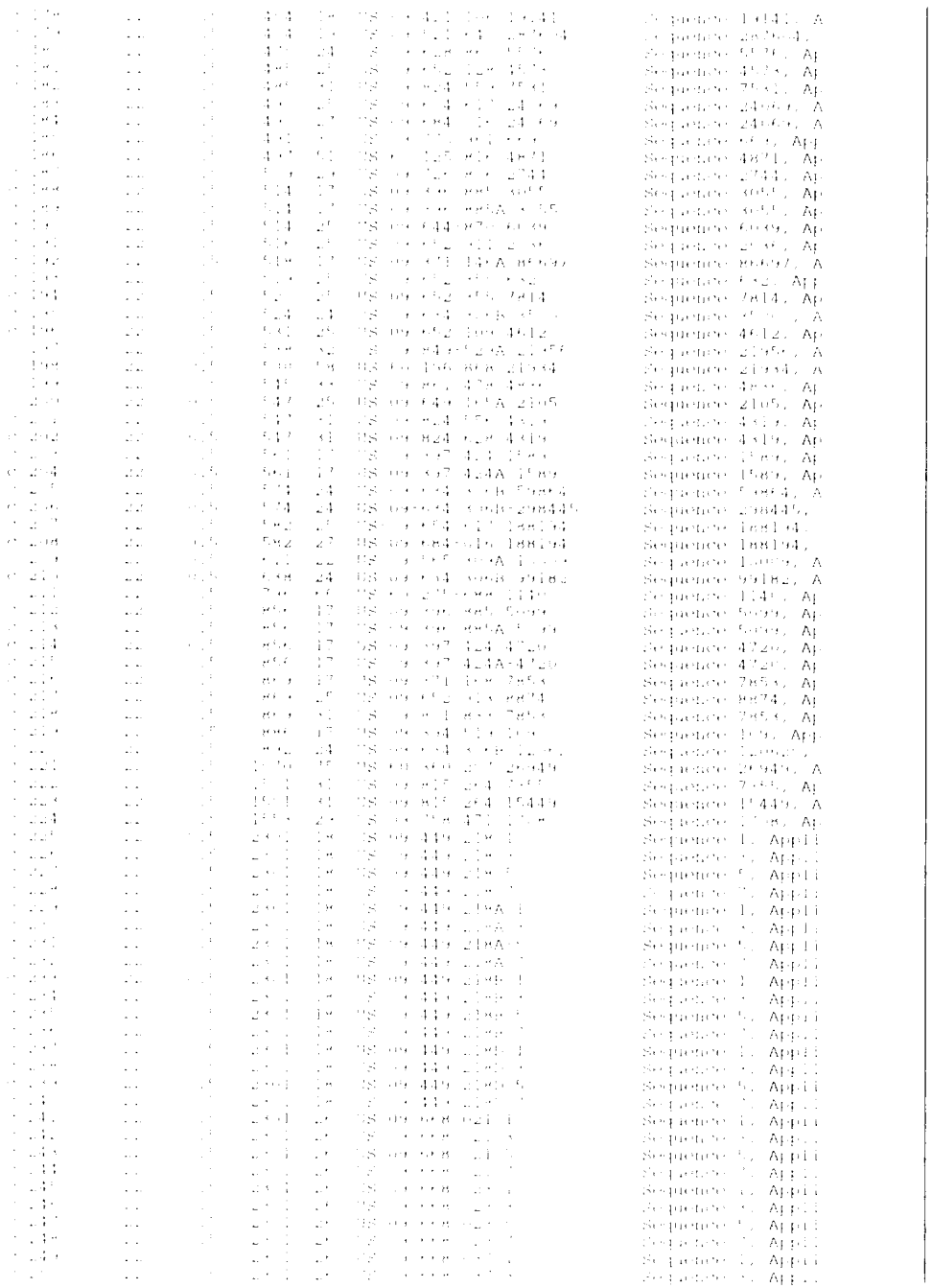


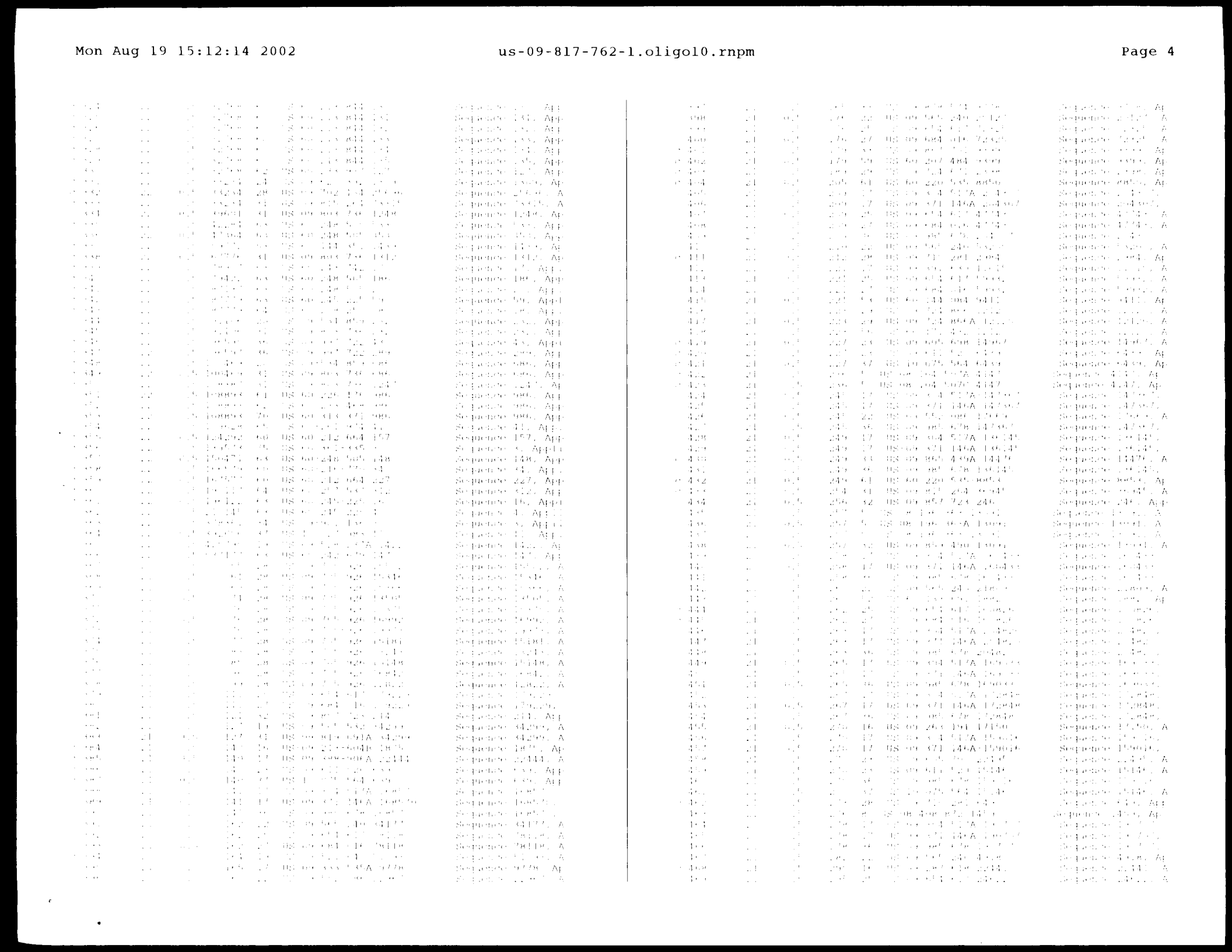


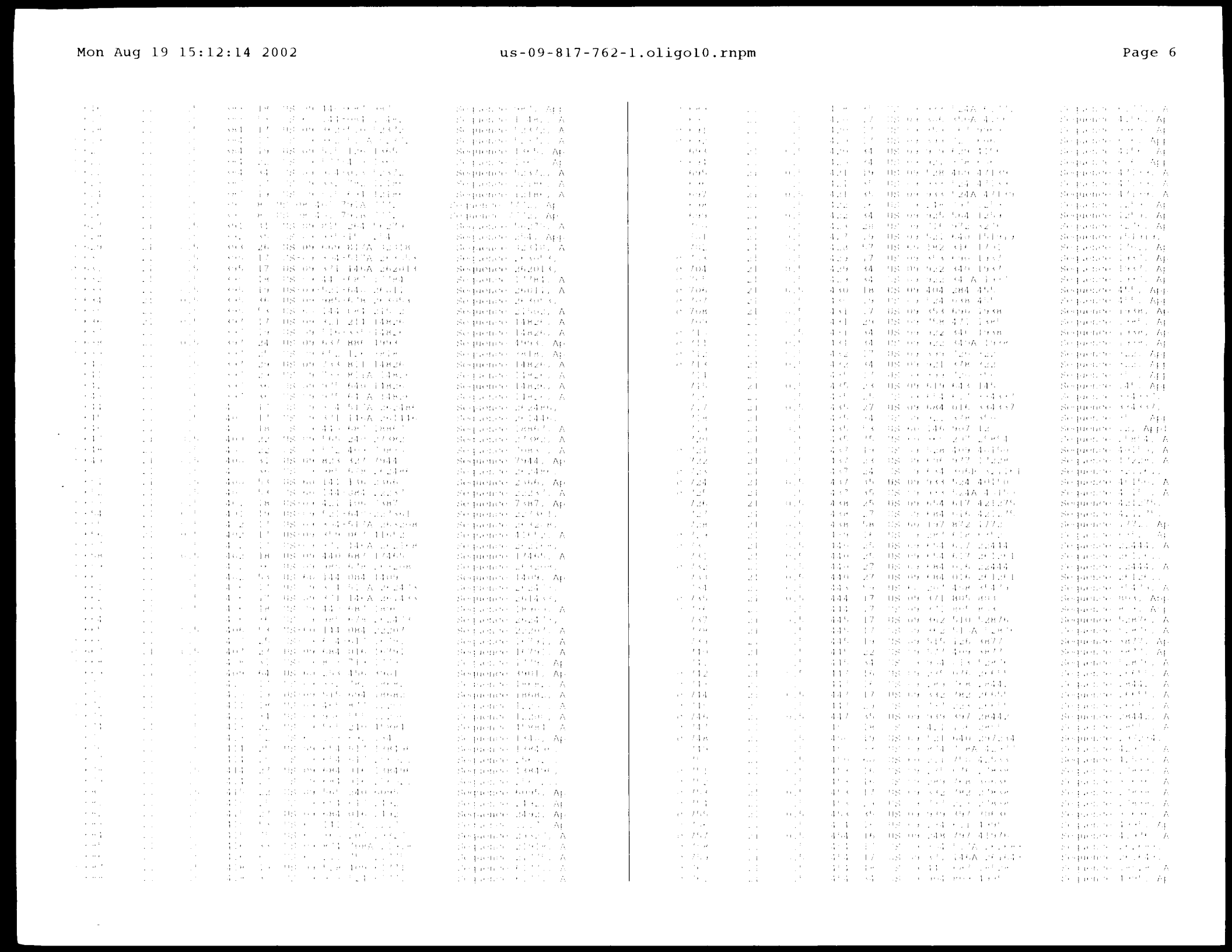


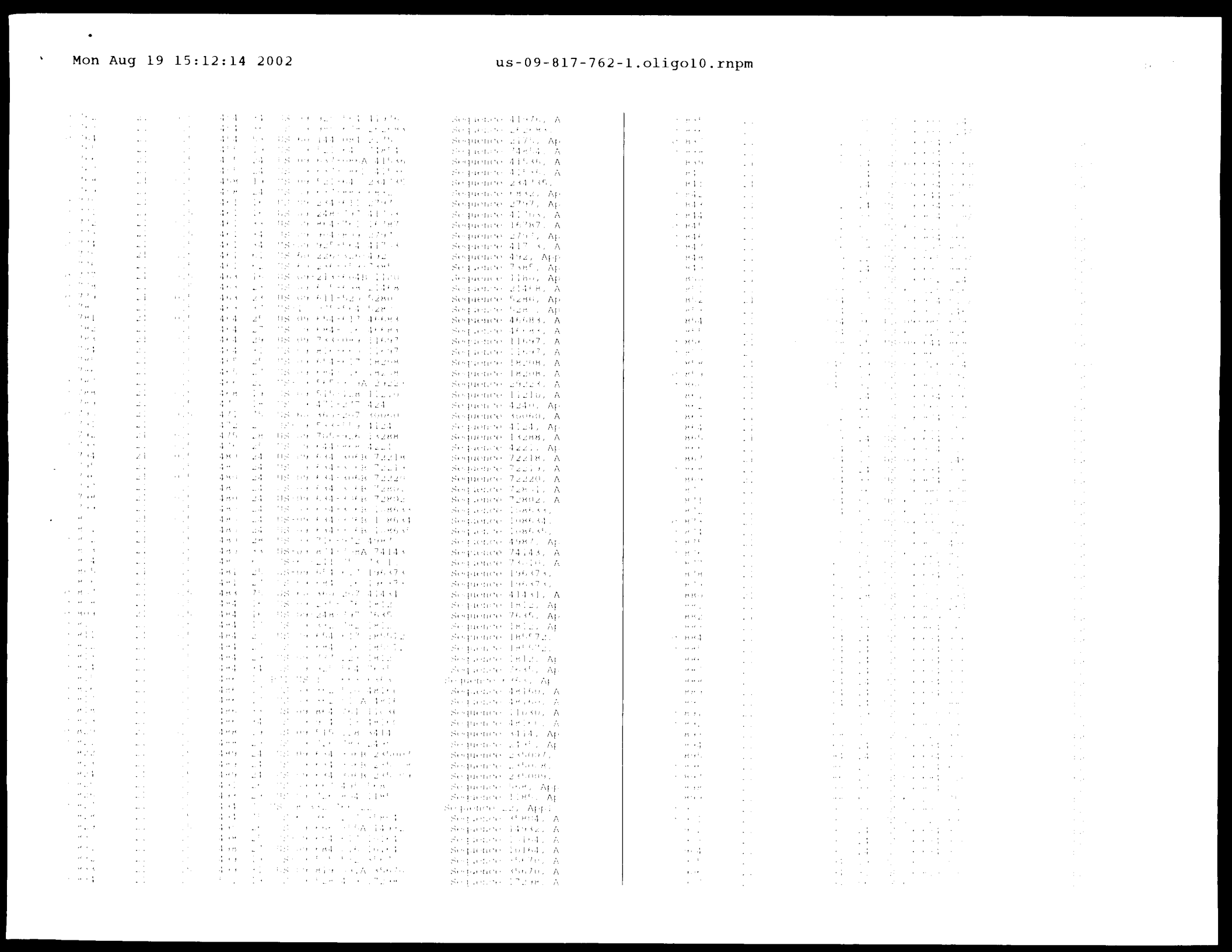


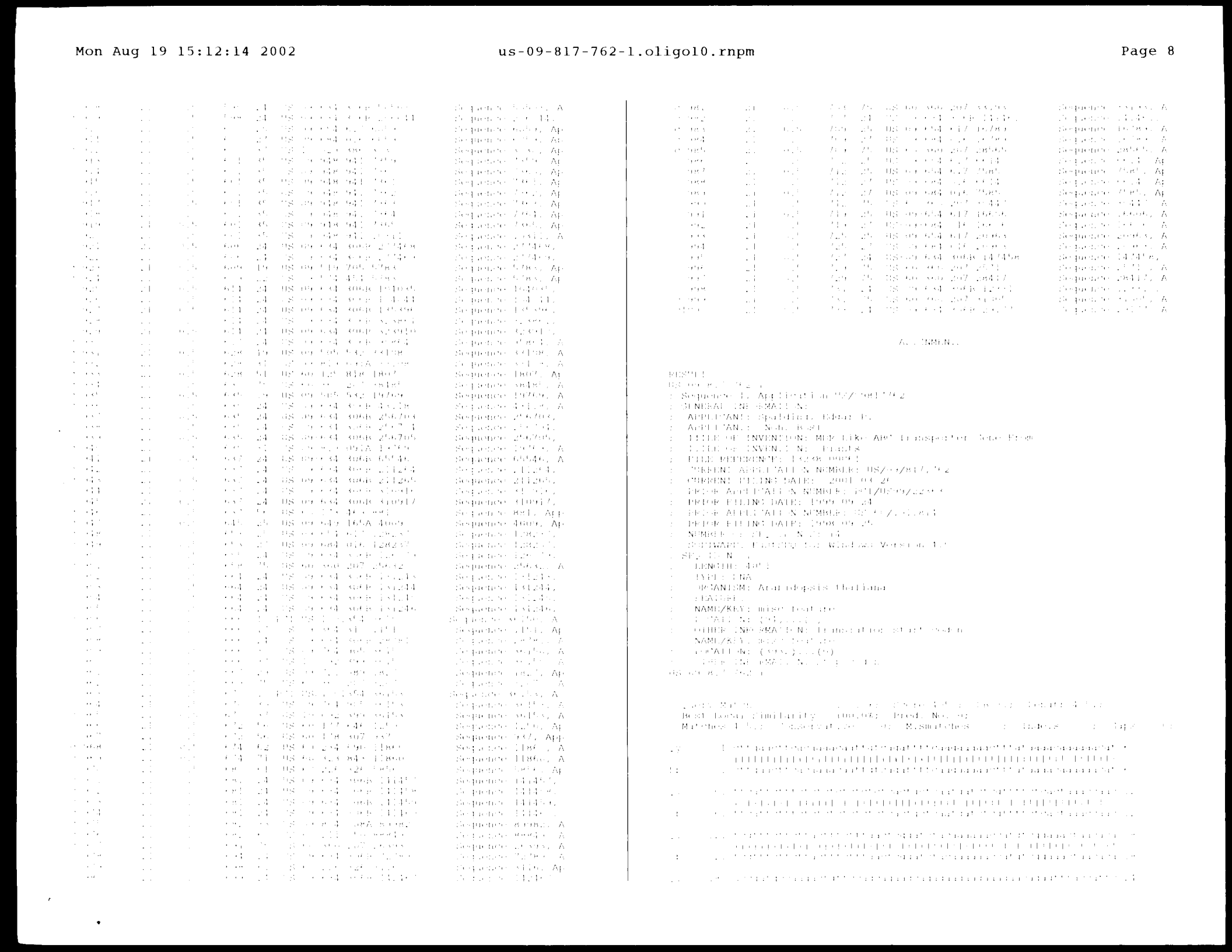


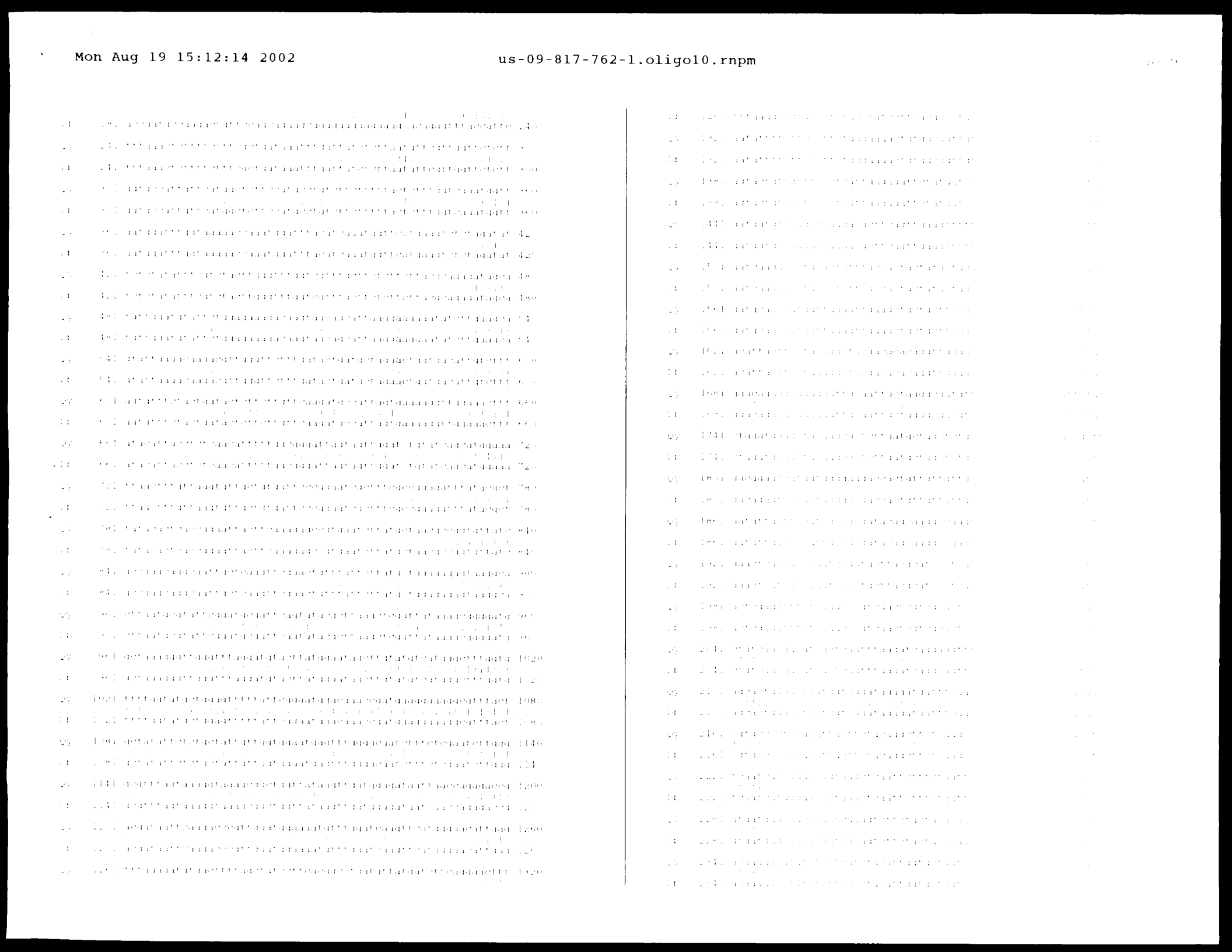


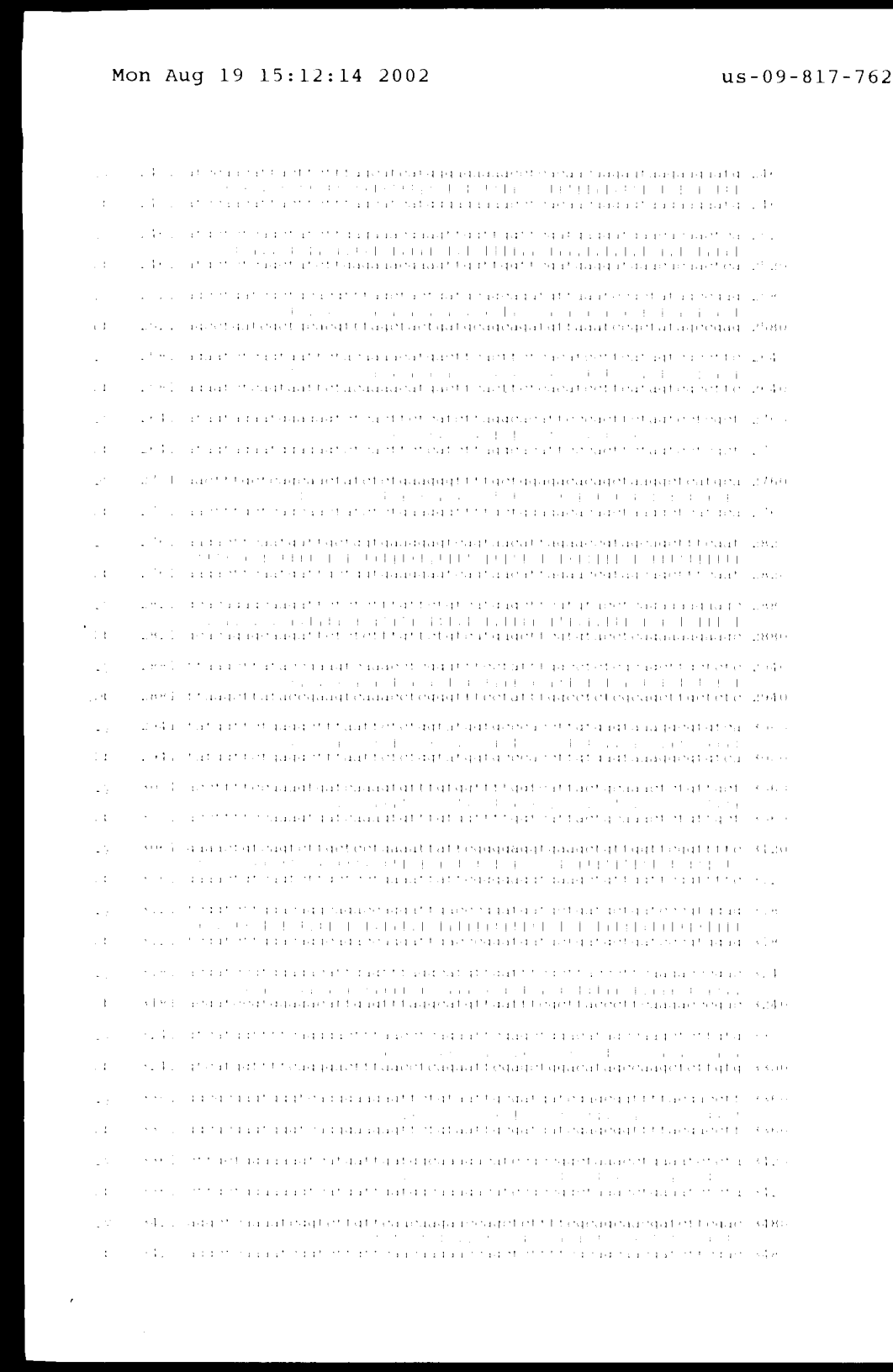


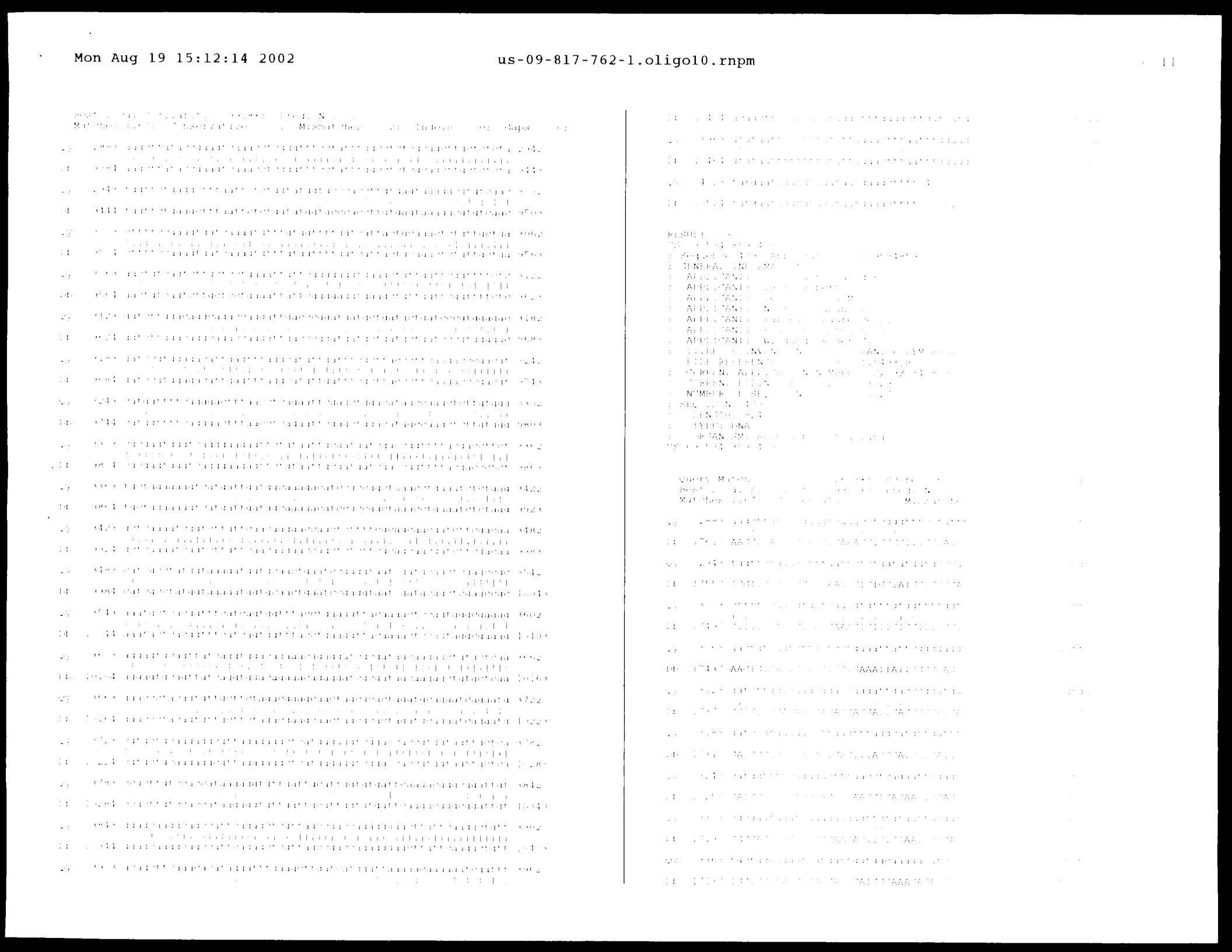


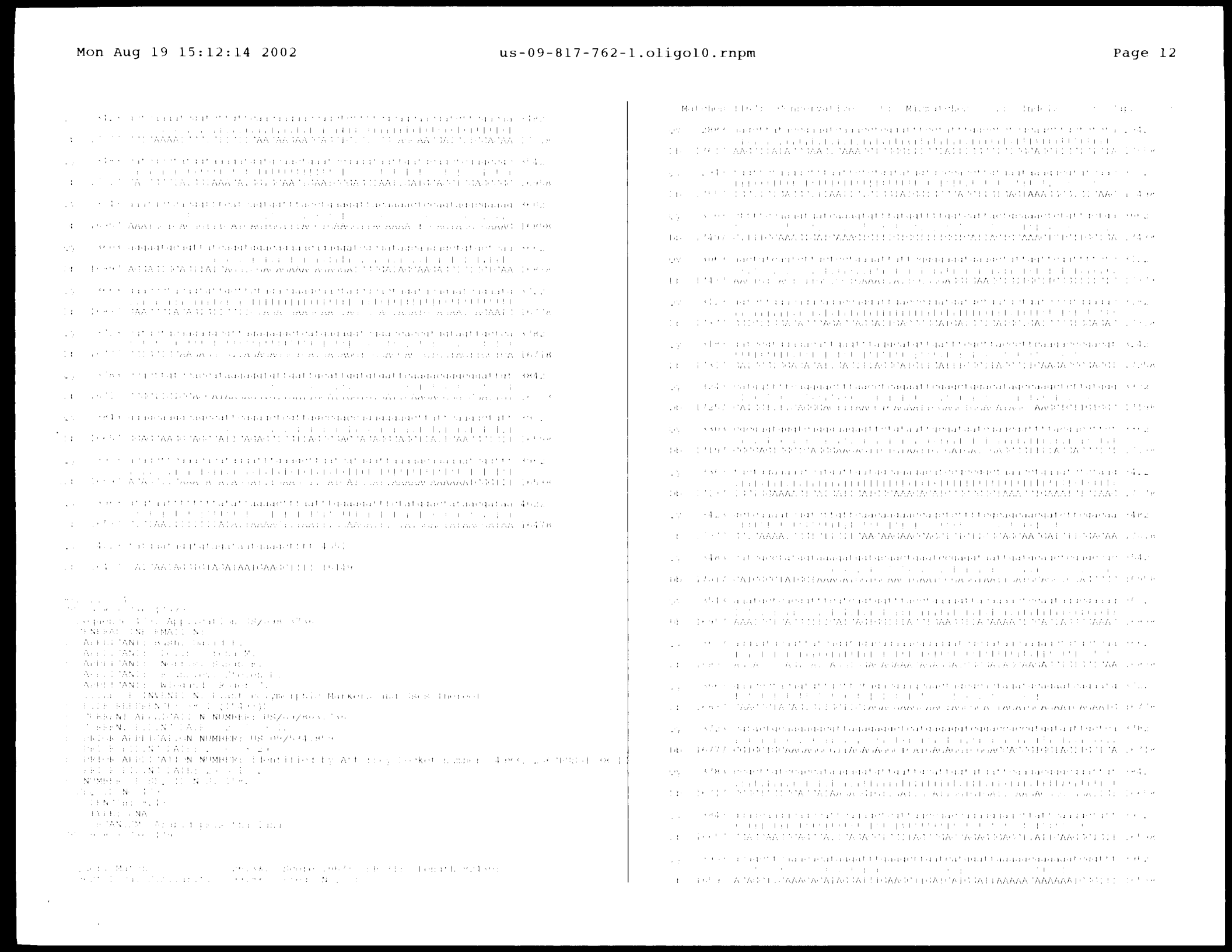


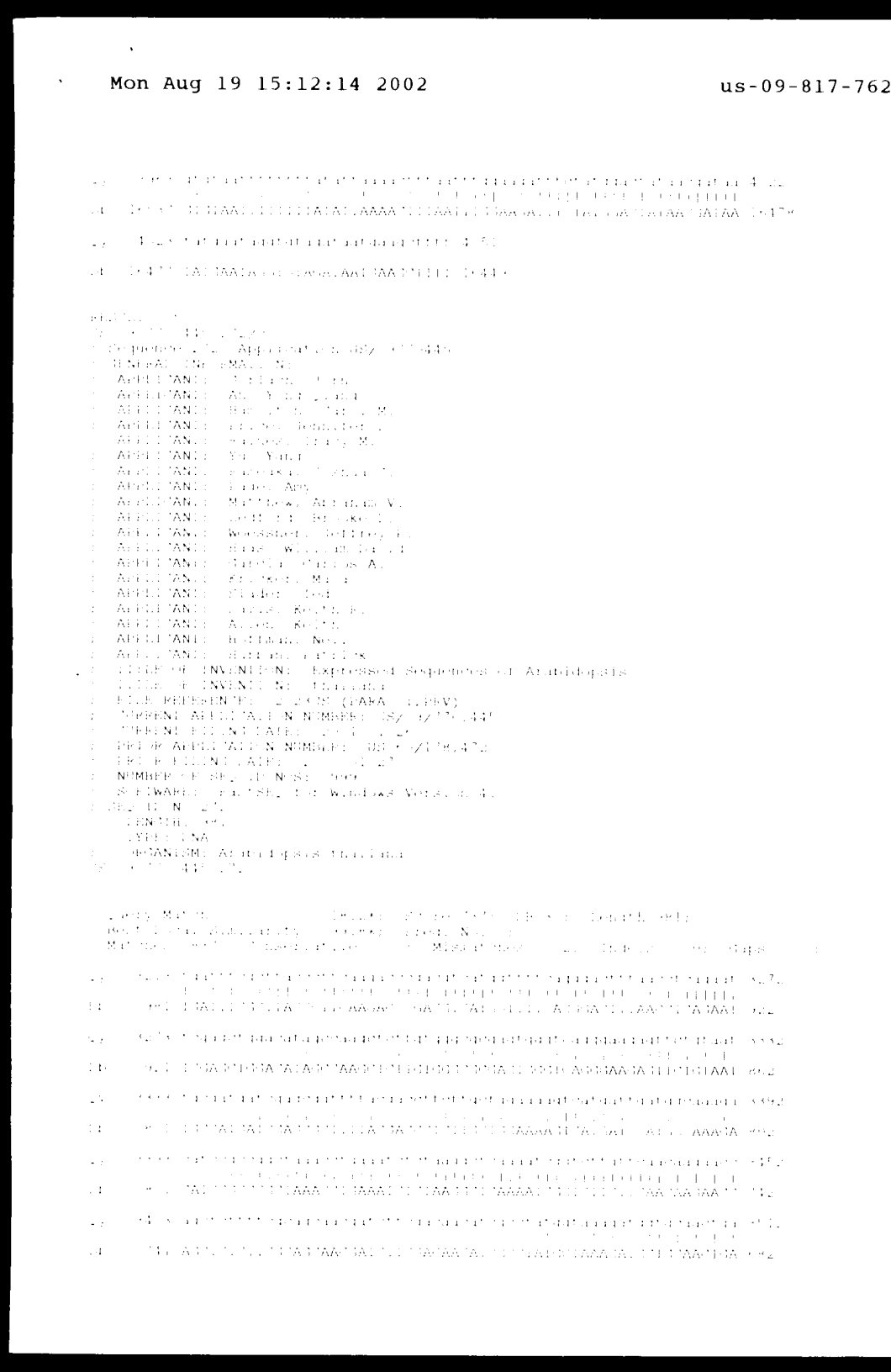


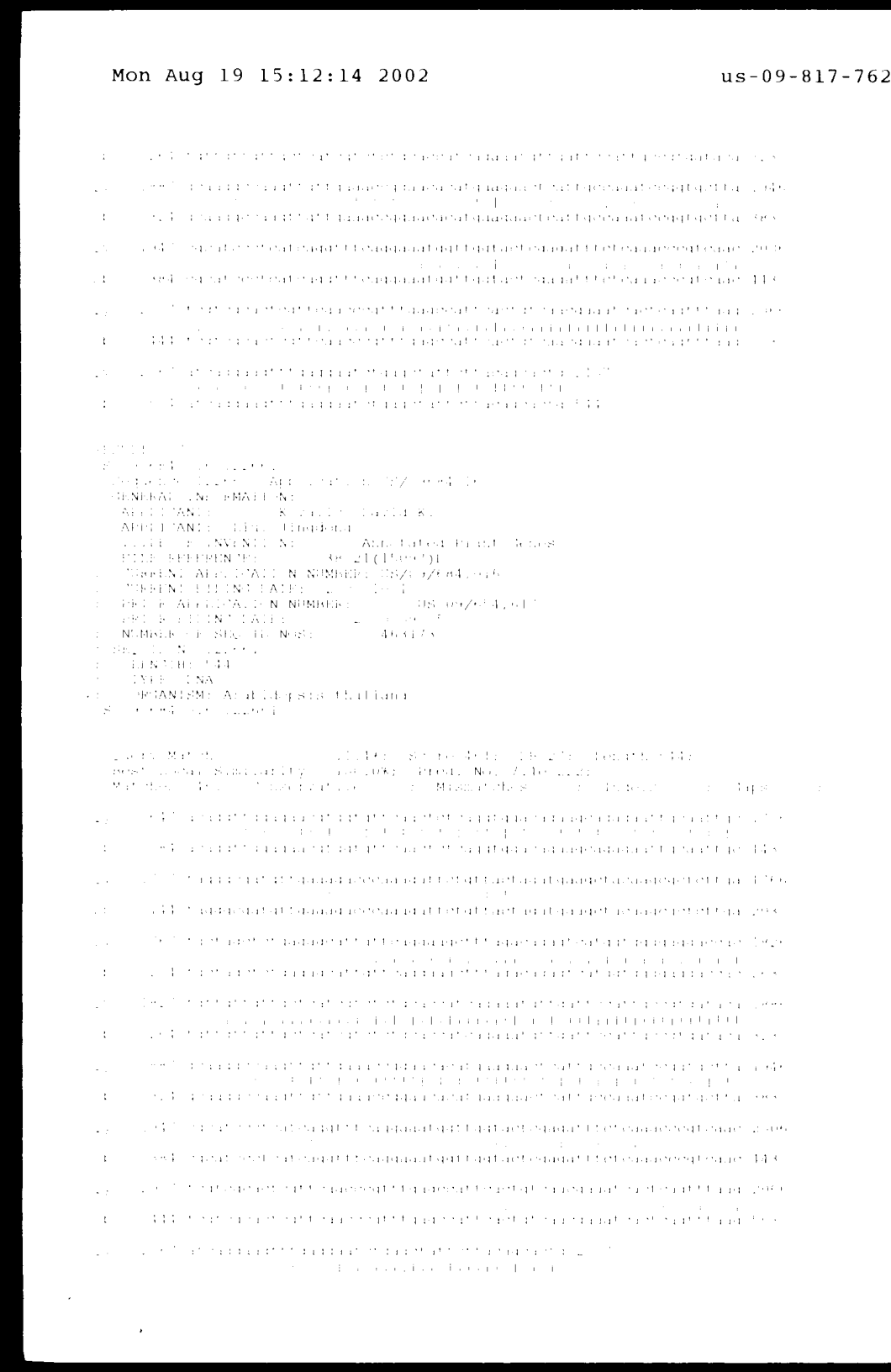


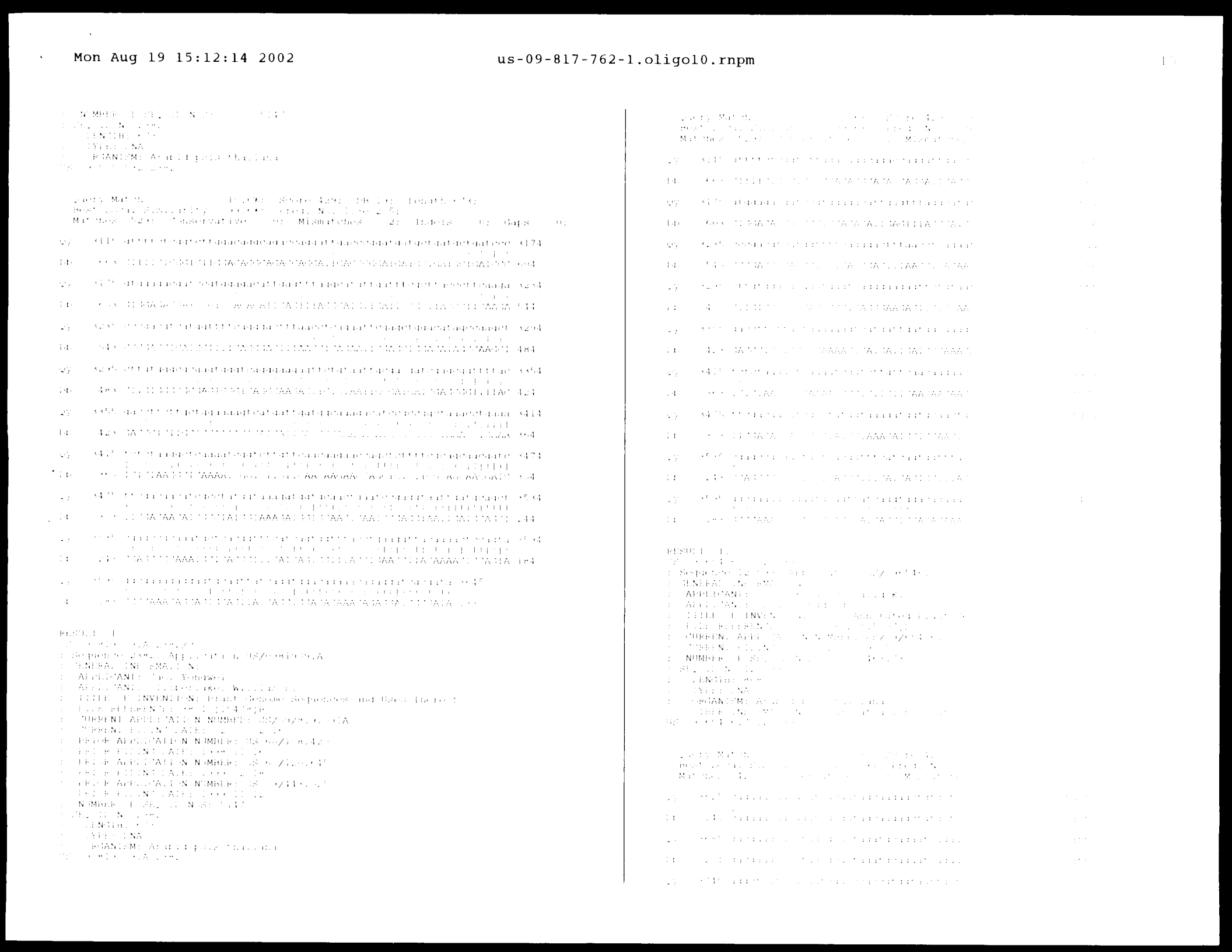


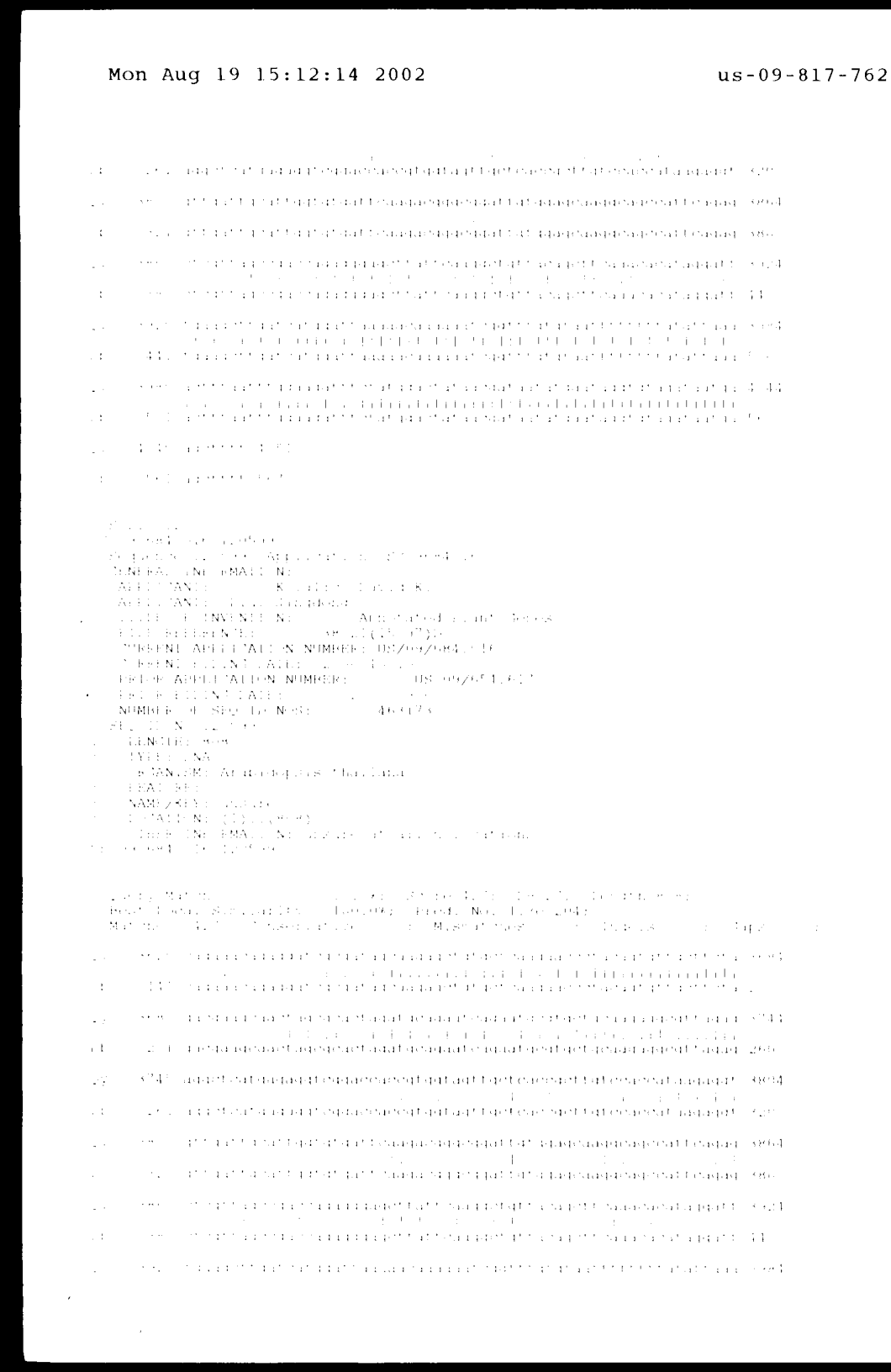


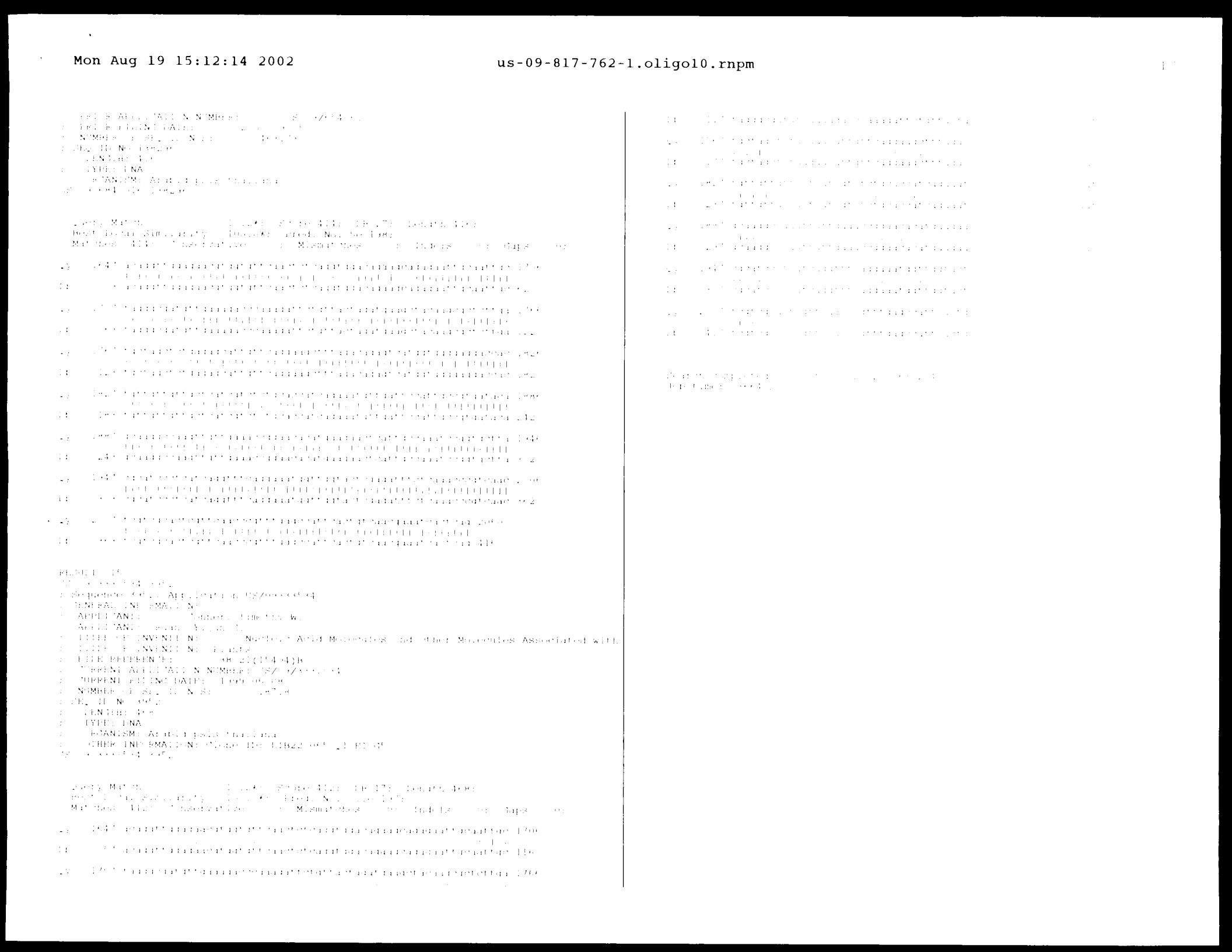


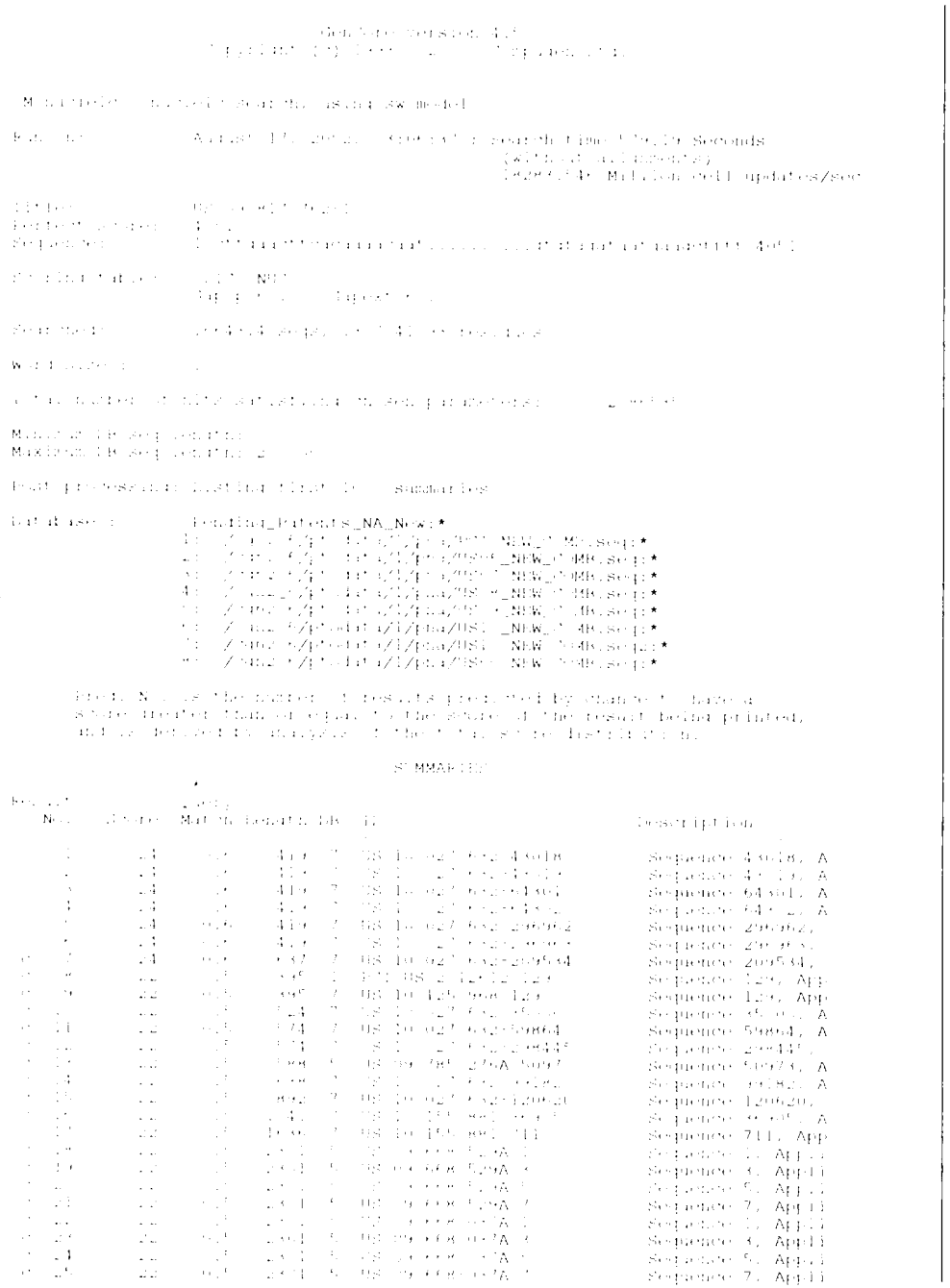


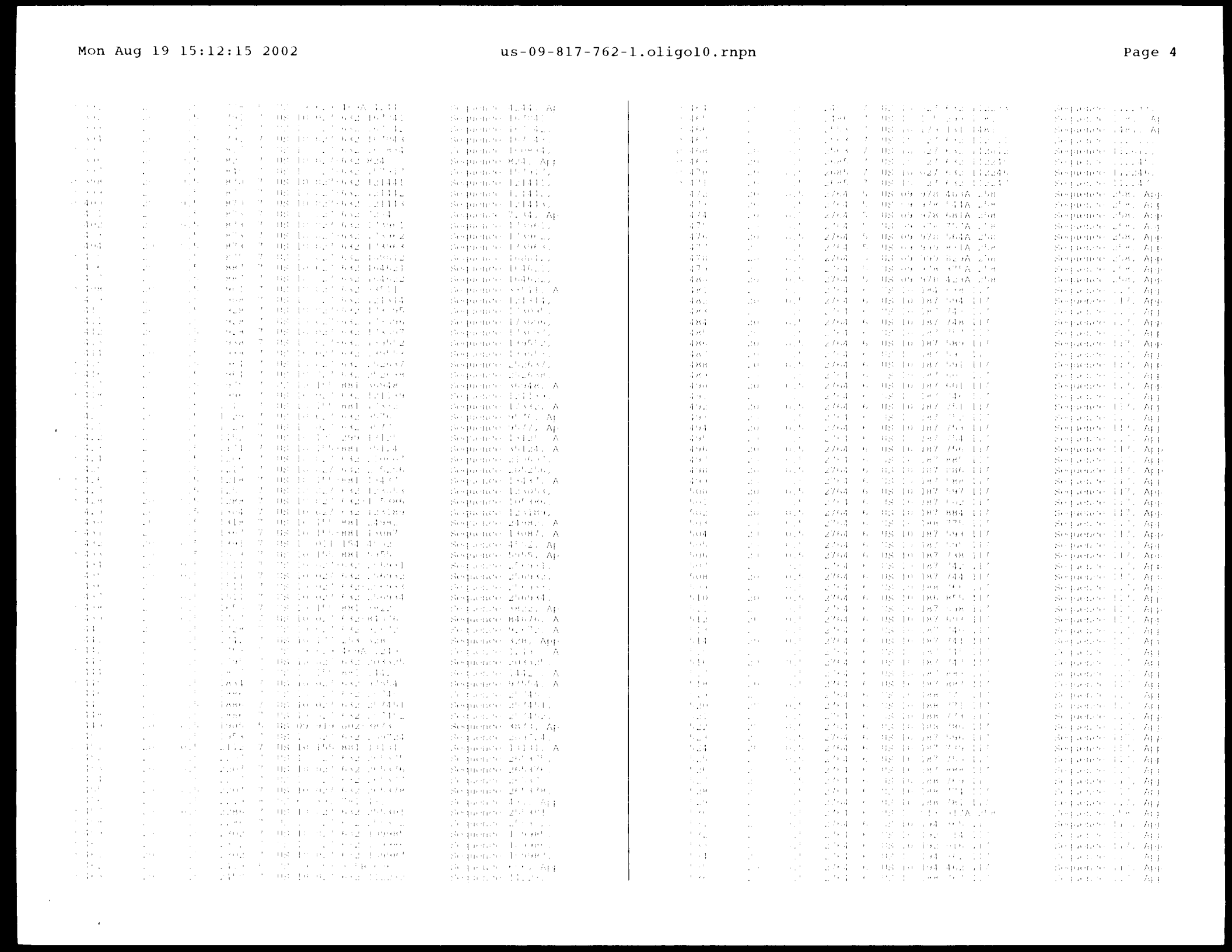


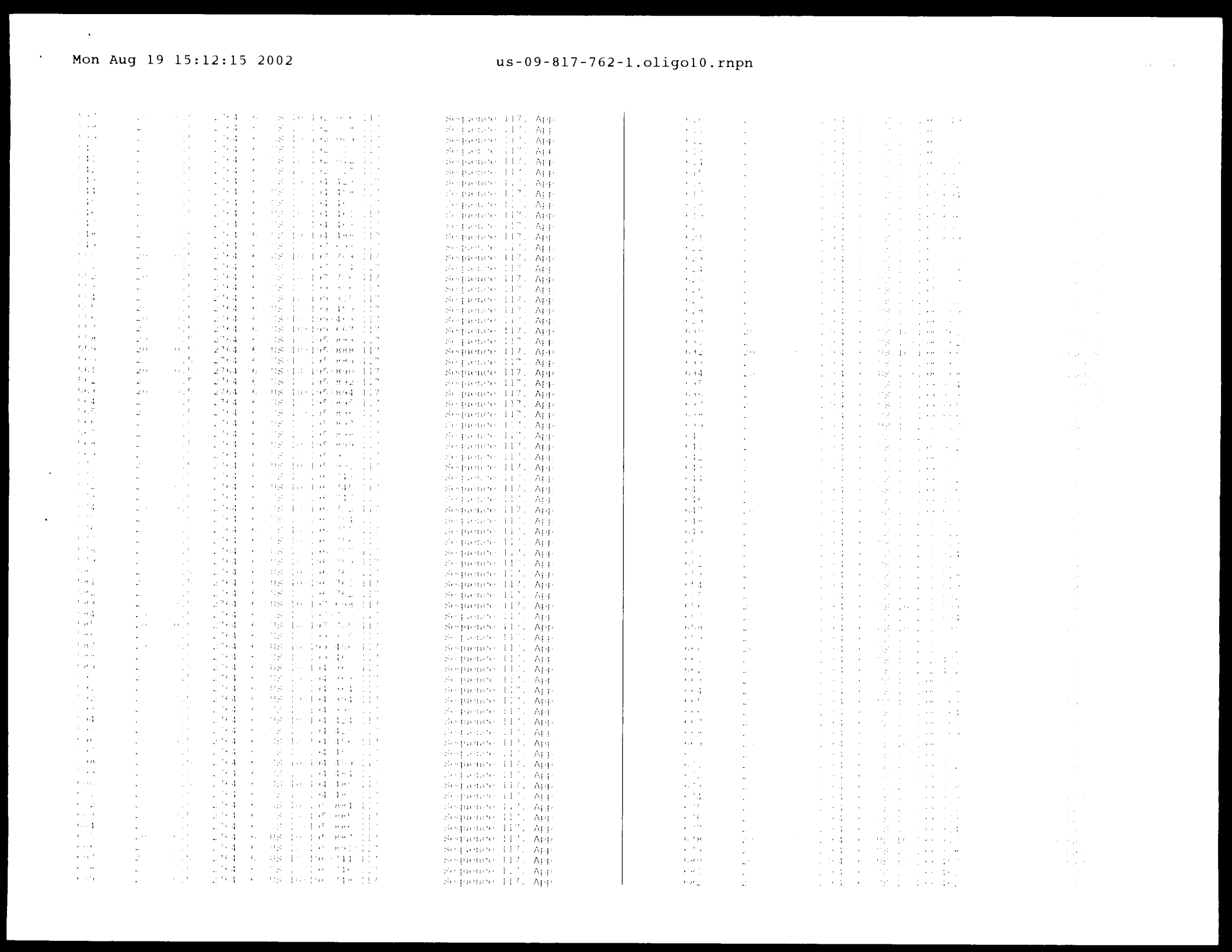




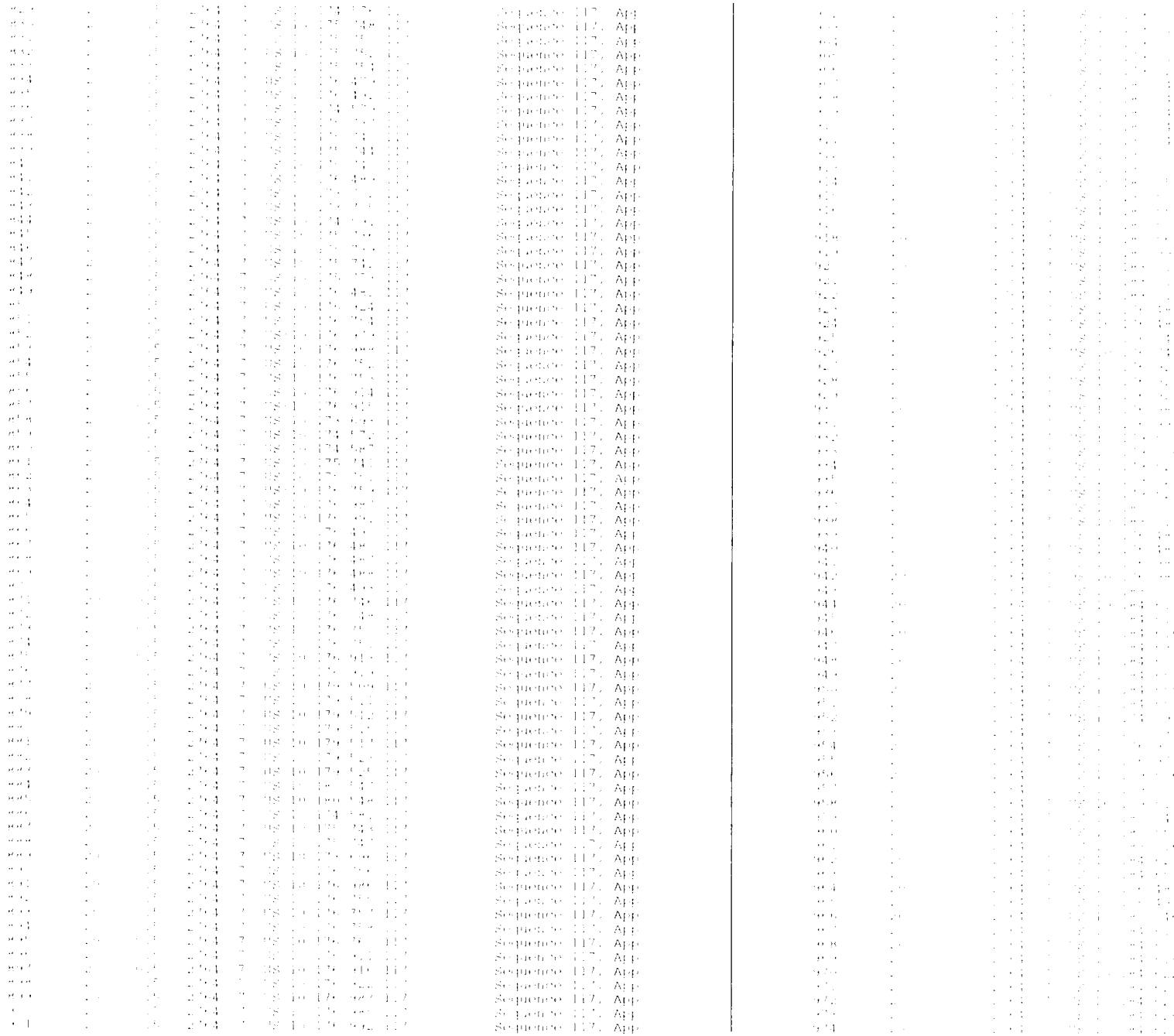


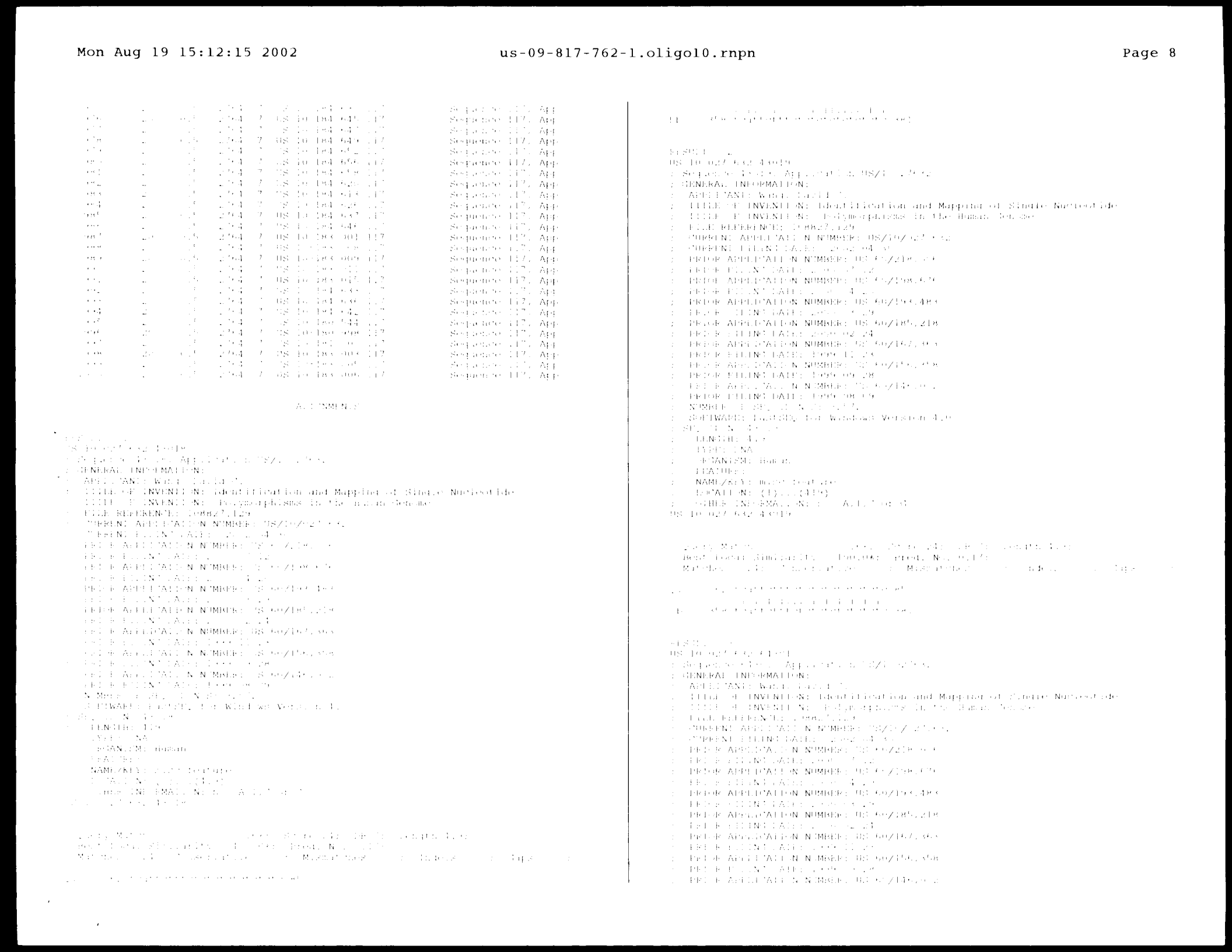


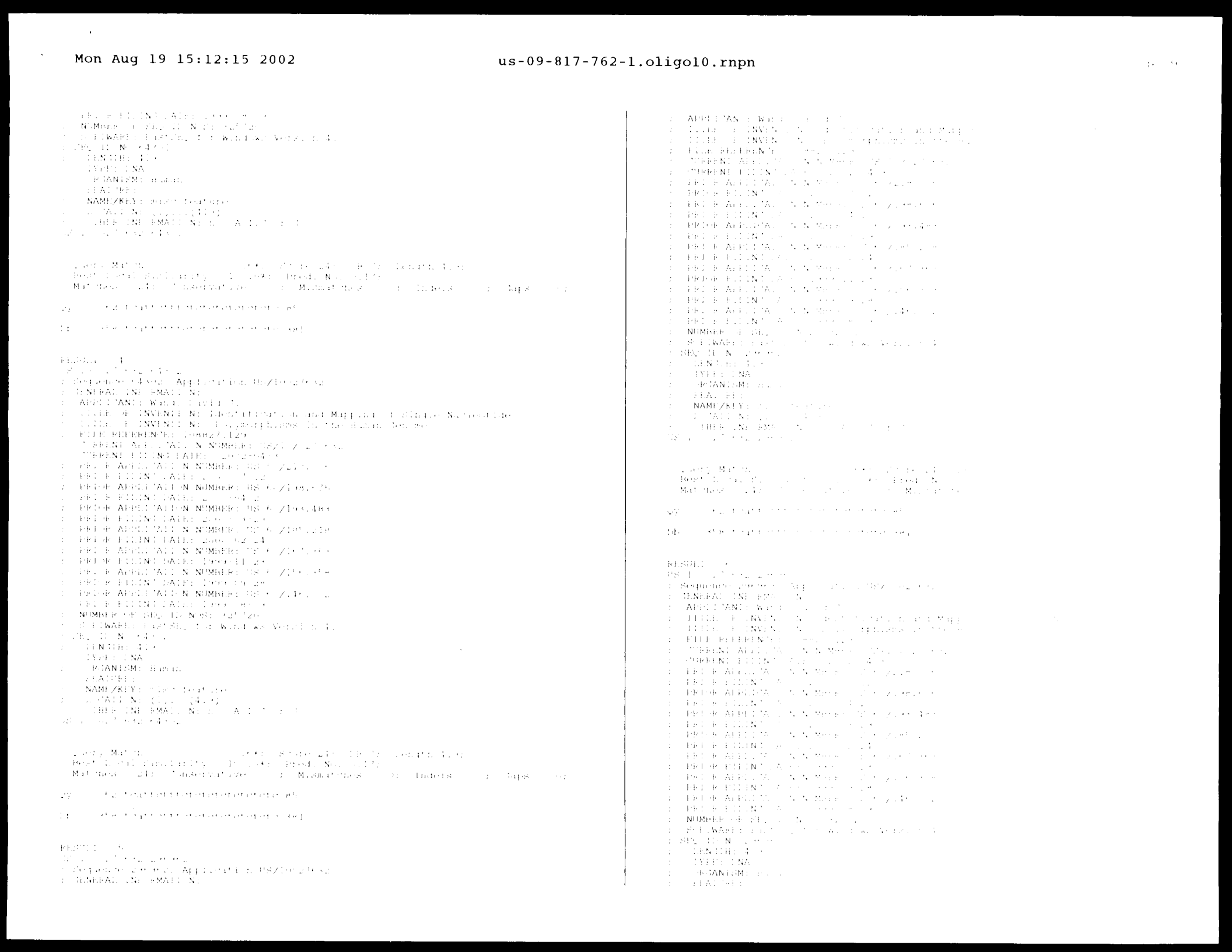


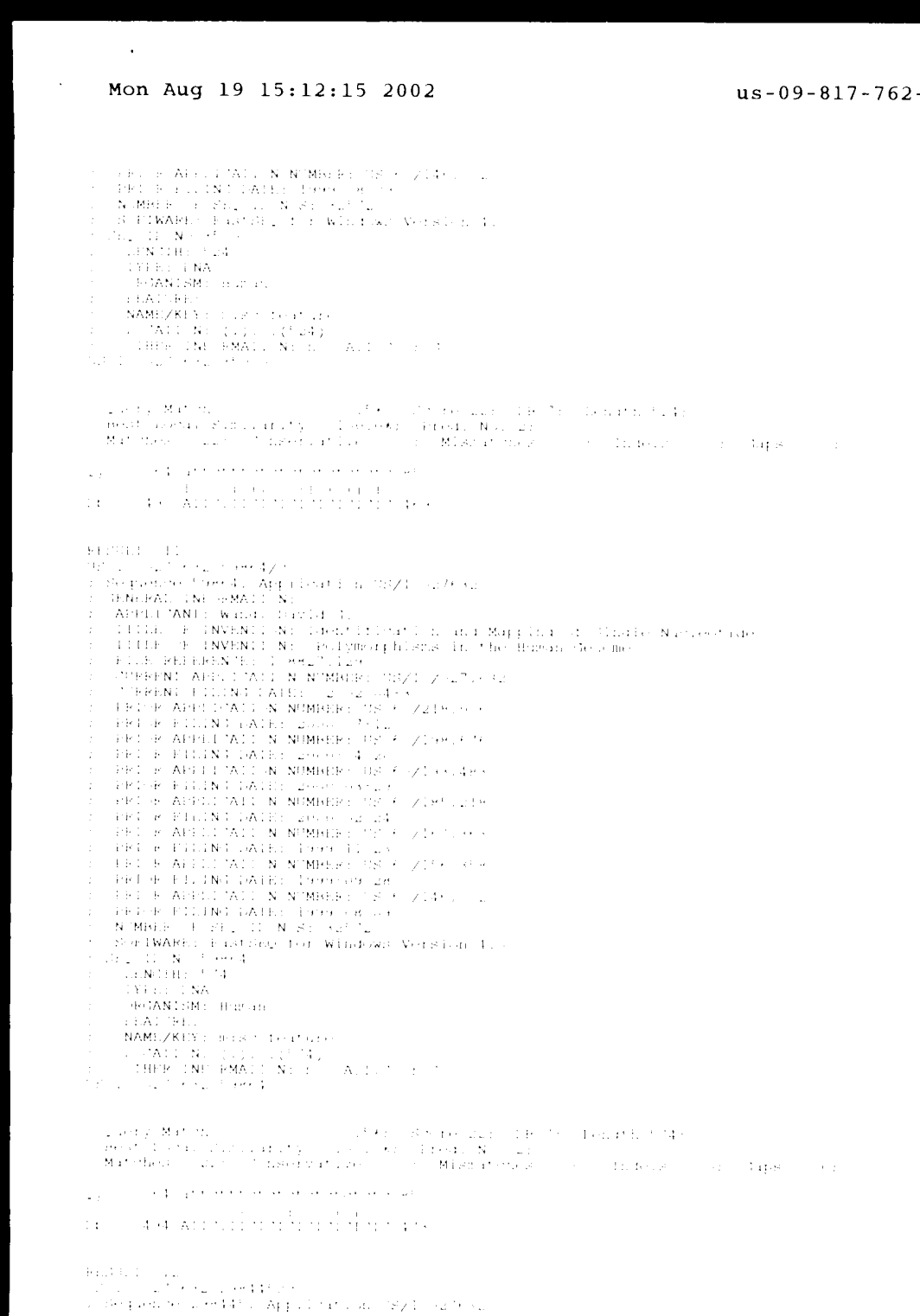


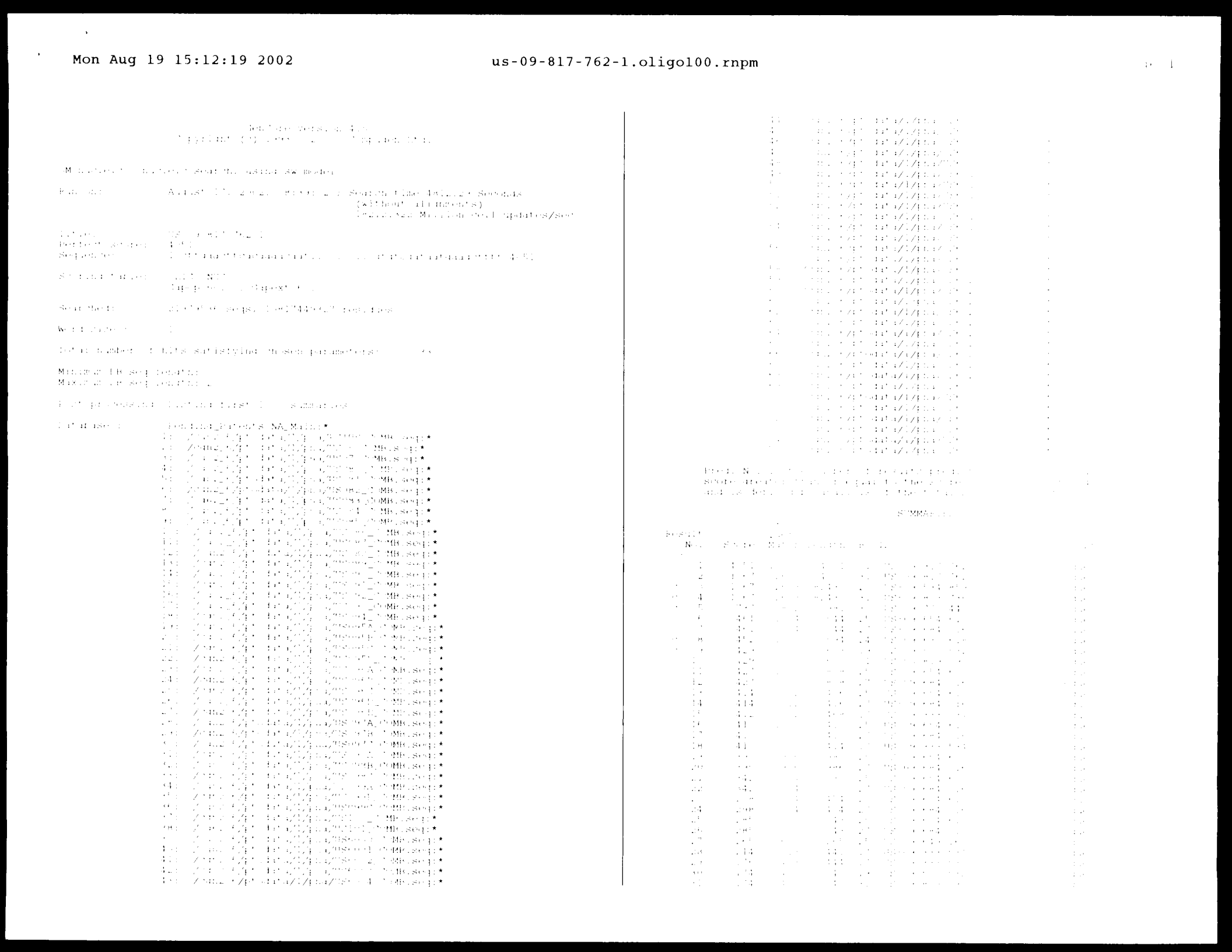




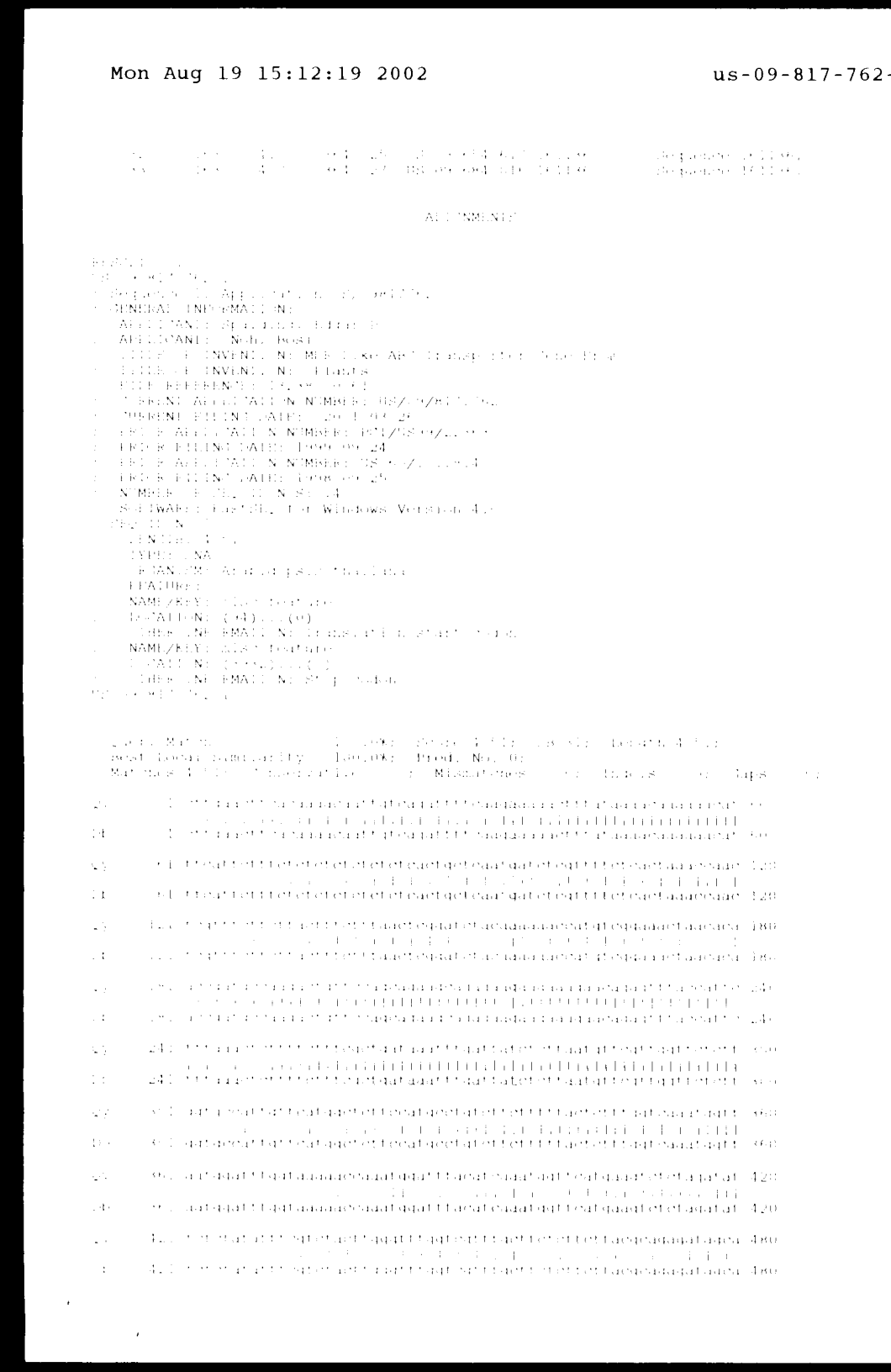


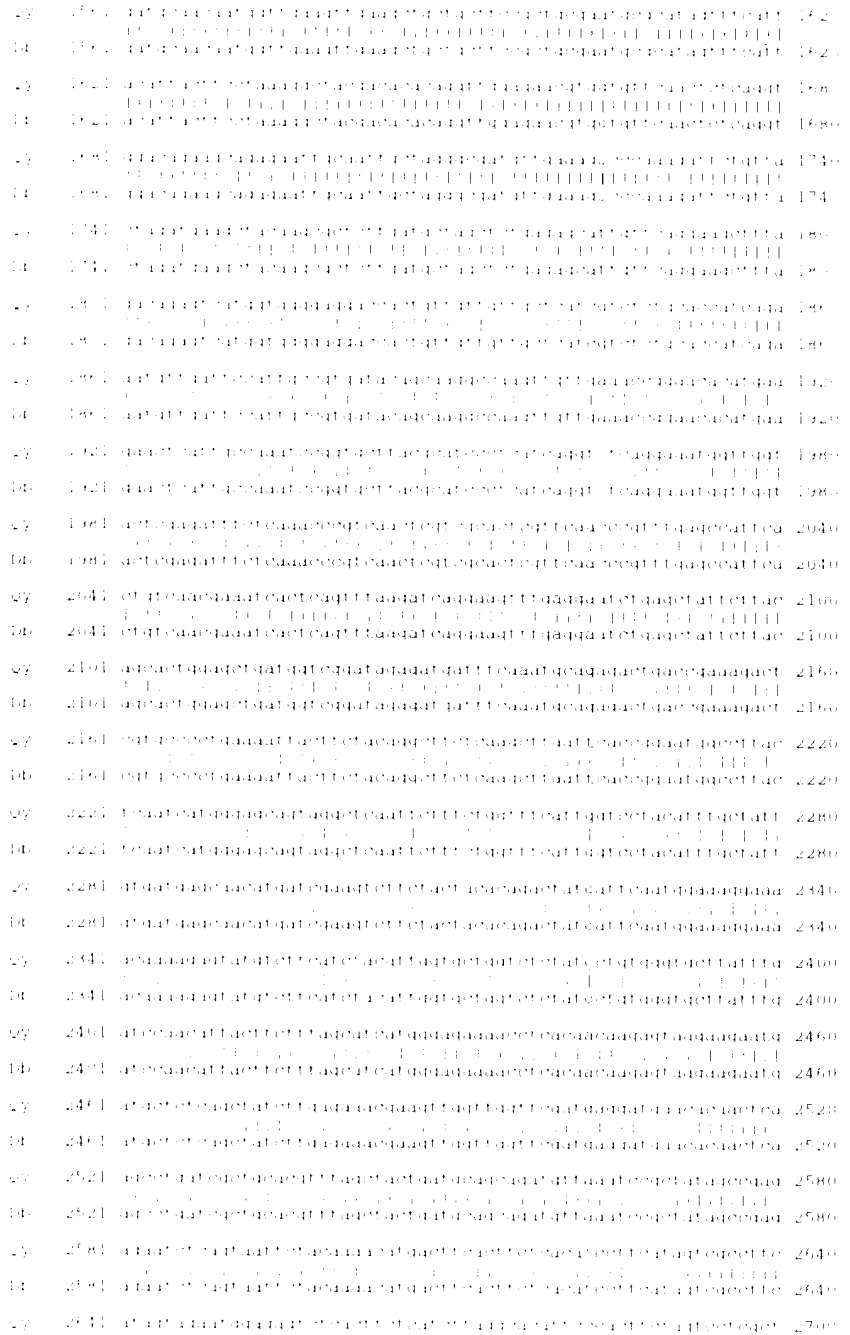


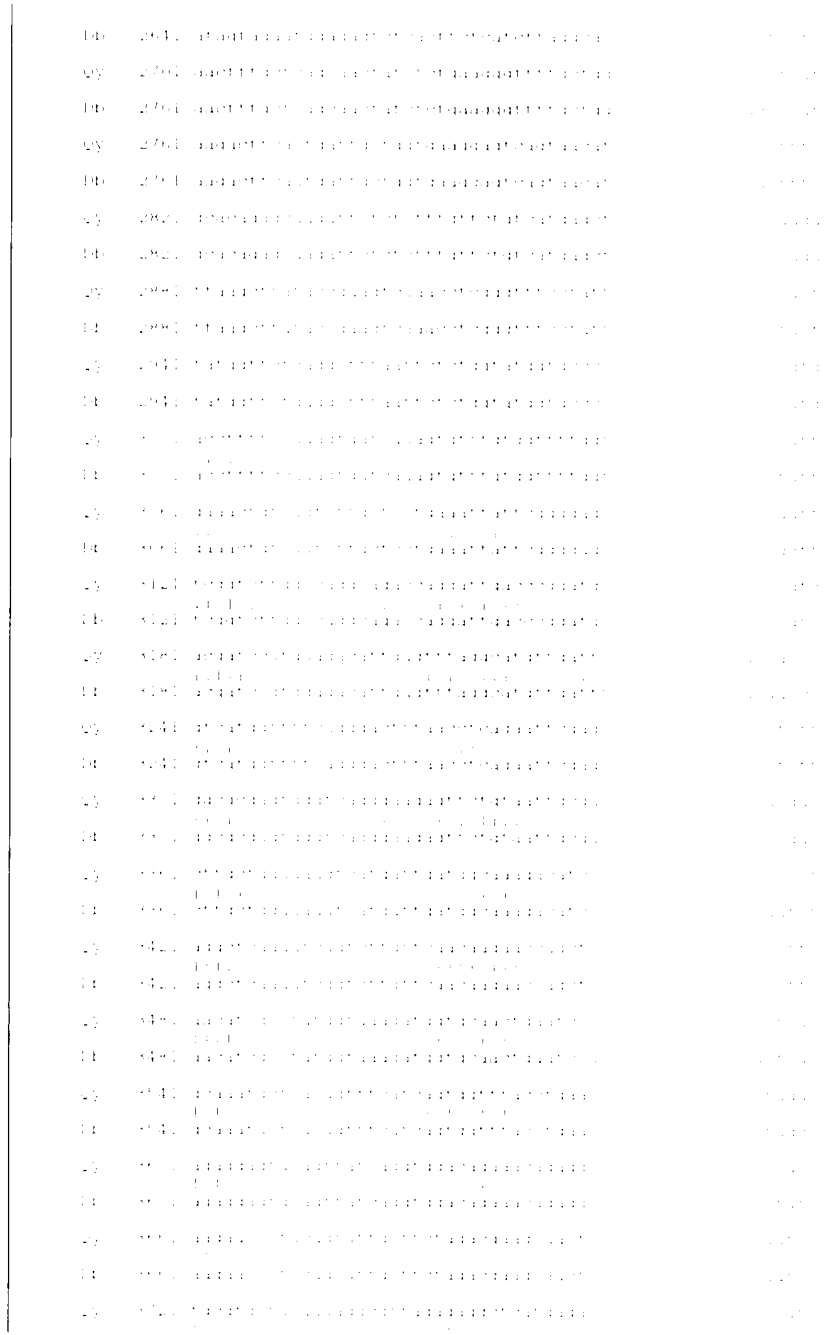


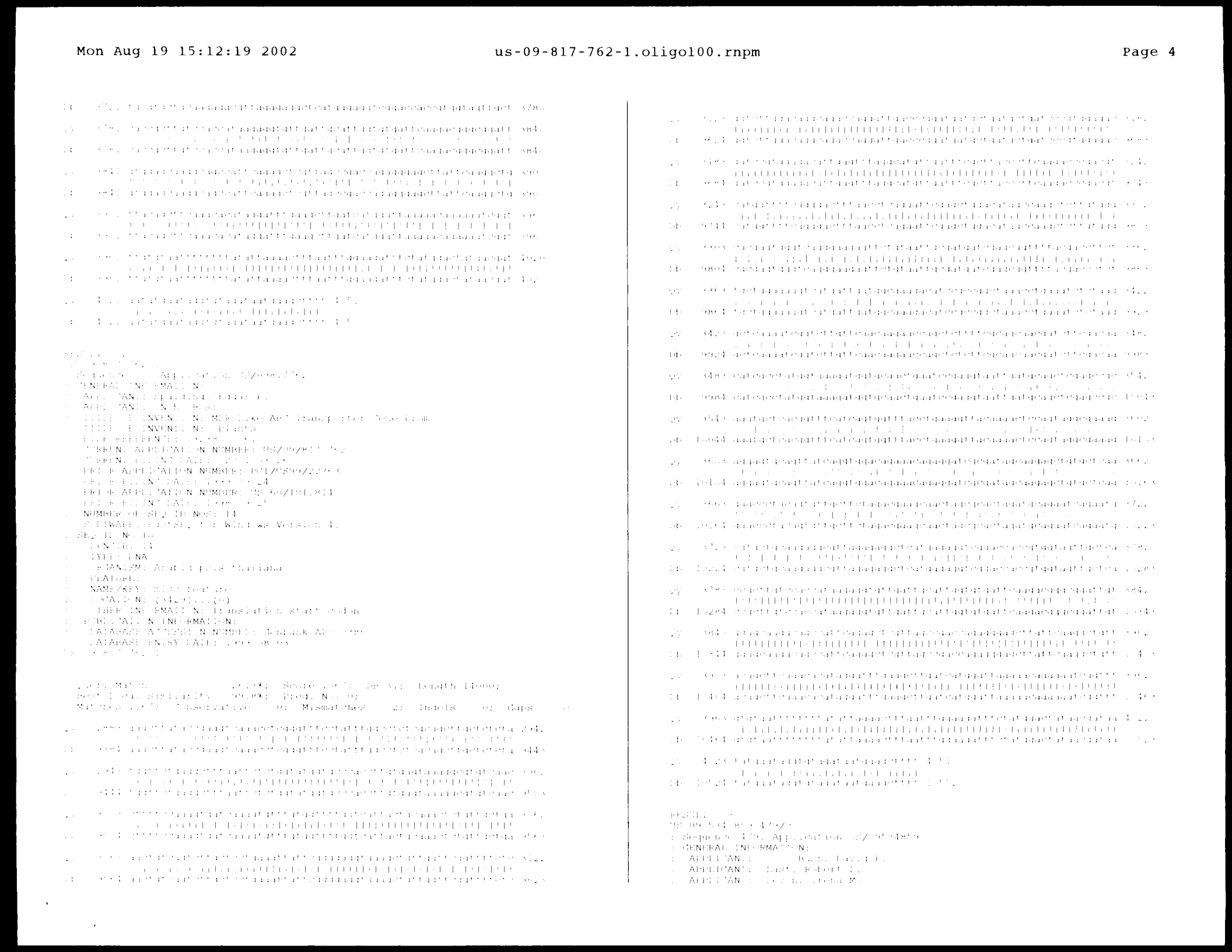


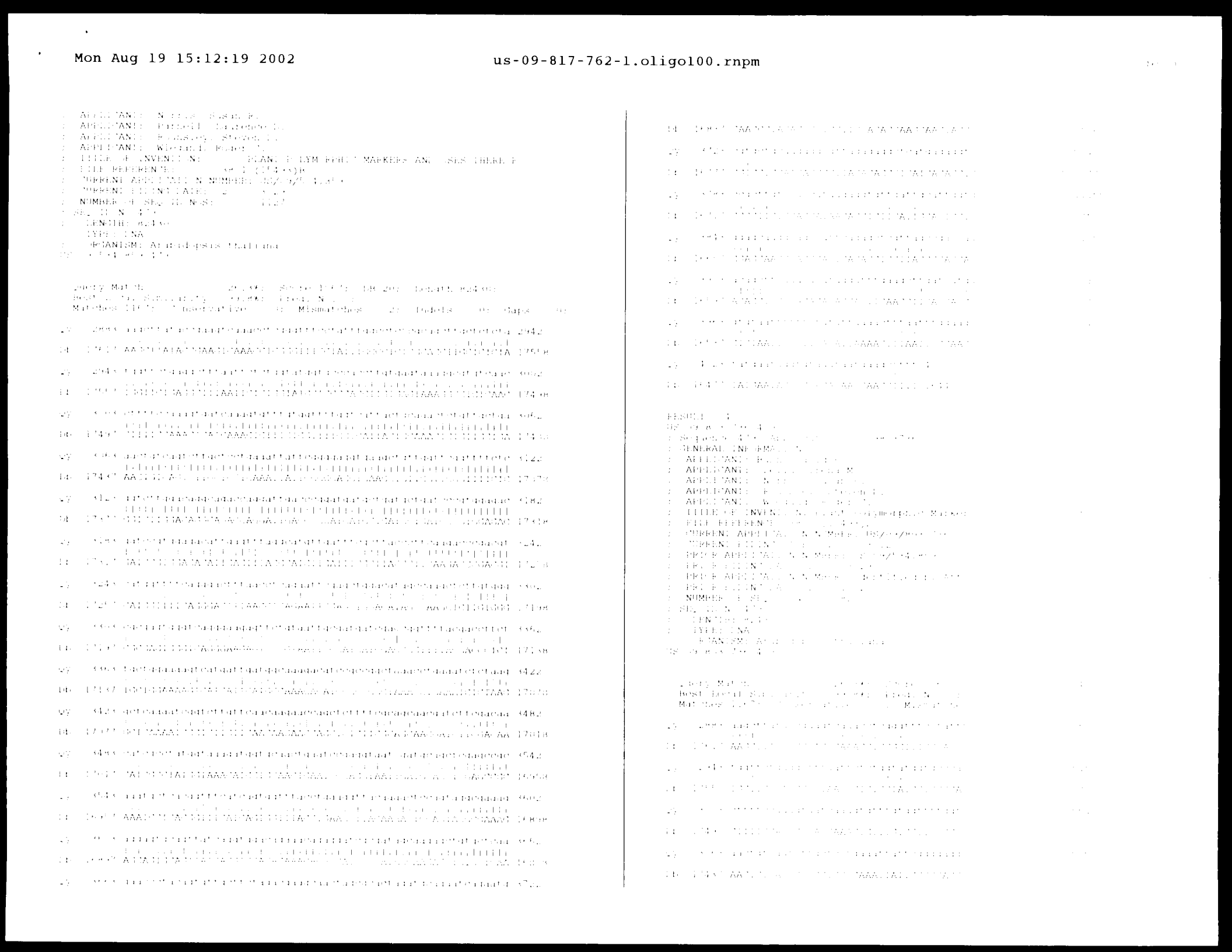
ALIGNMENTS

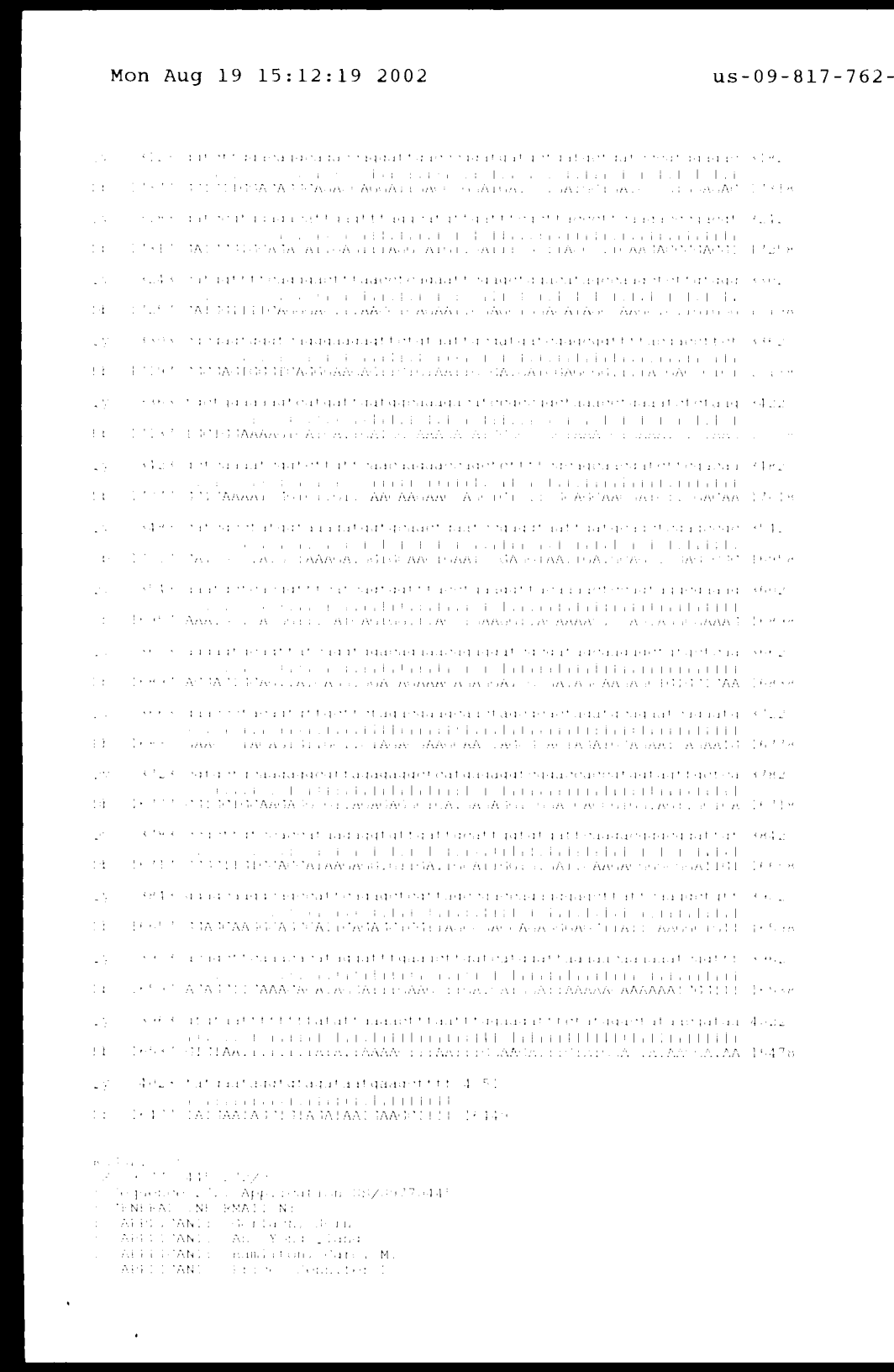


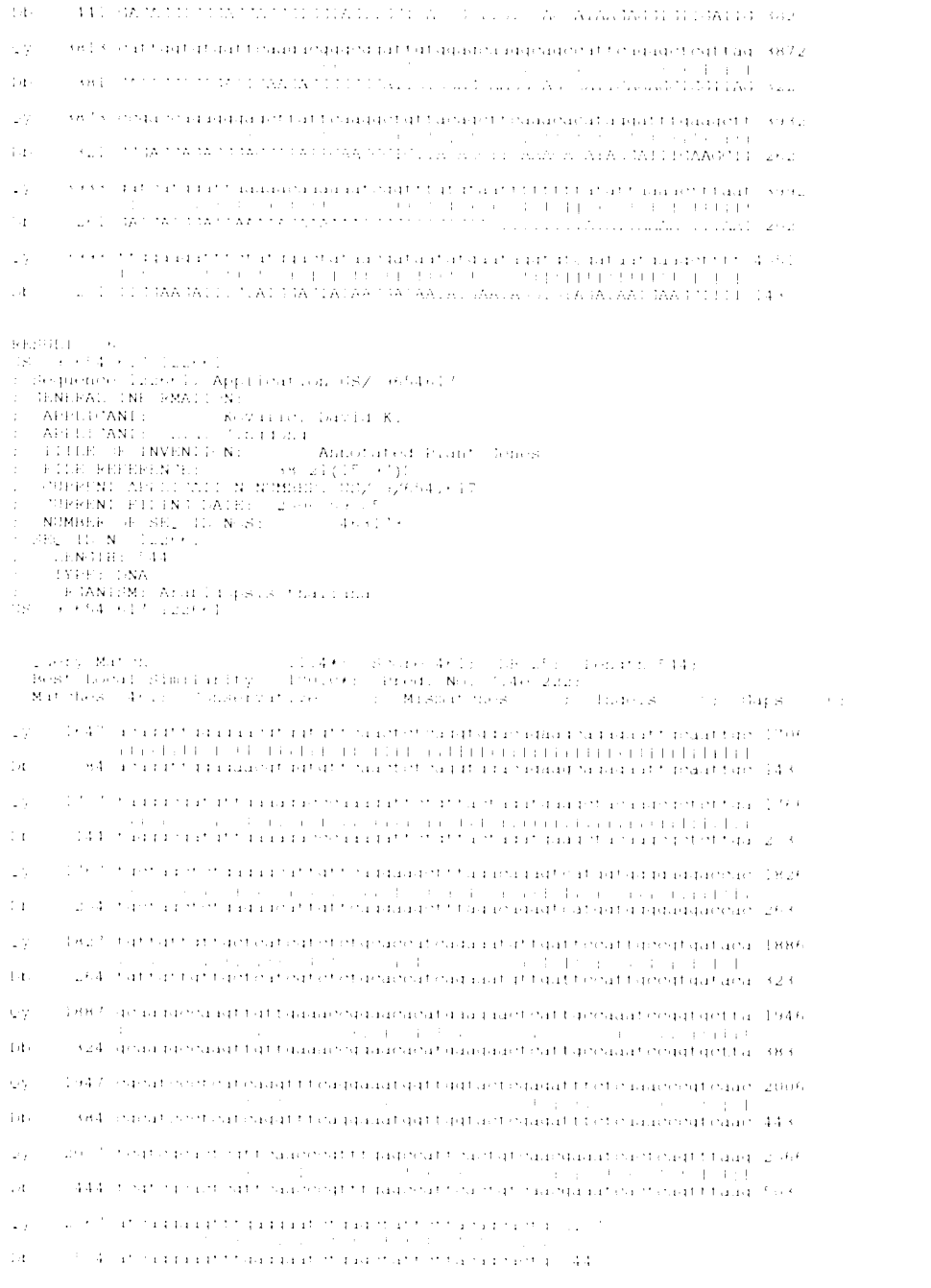


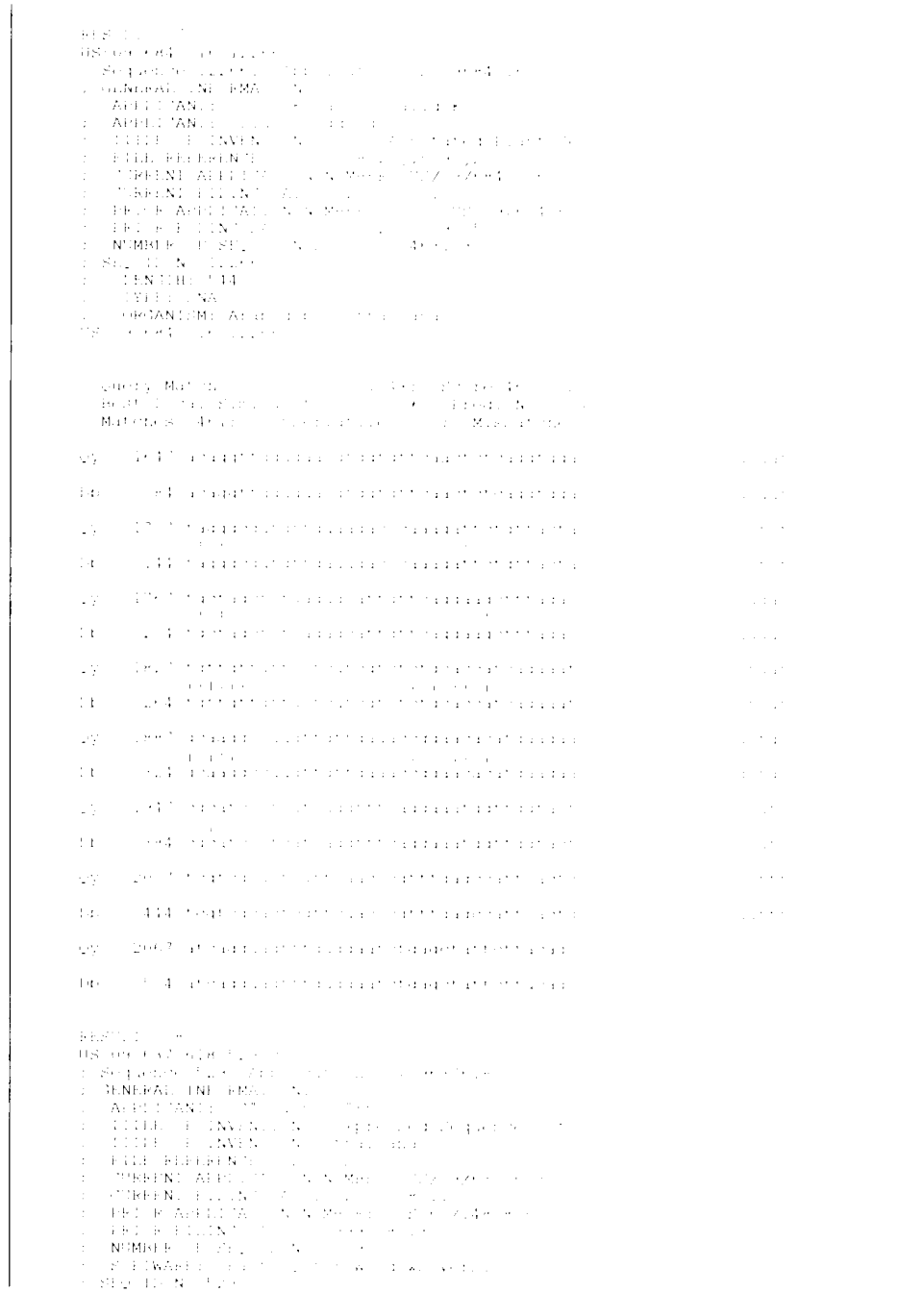


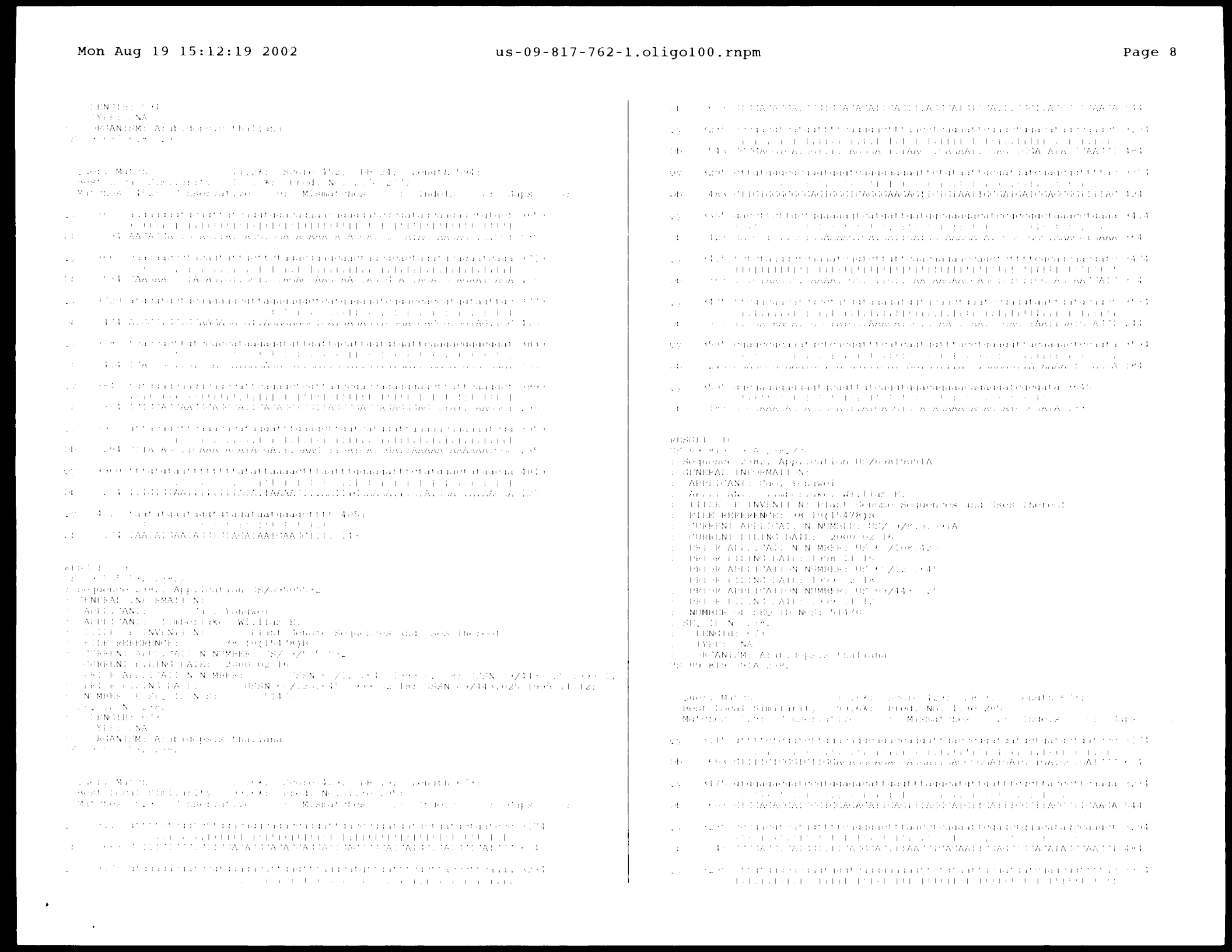


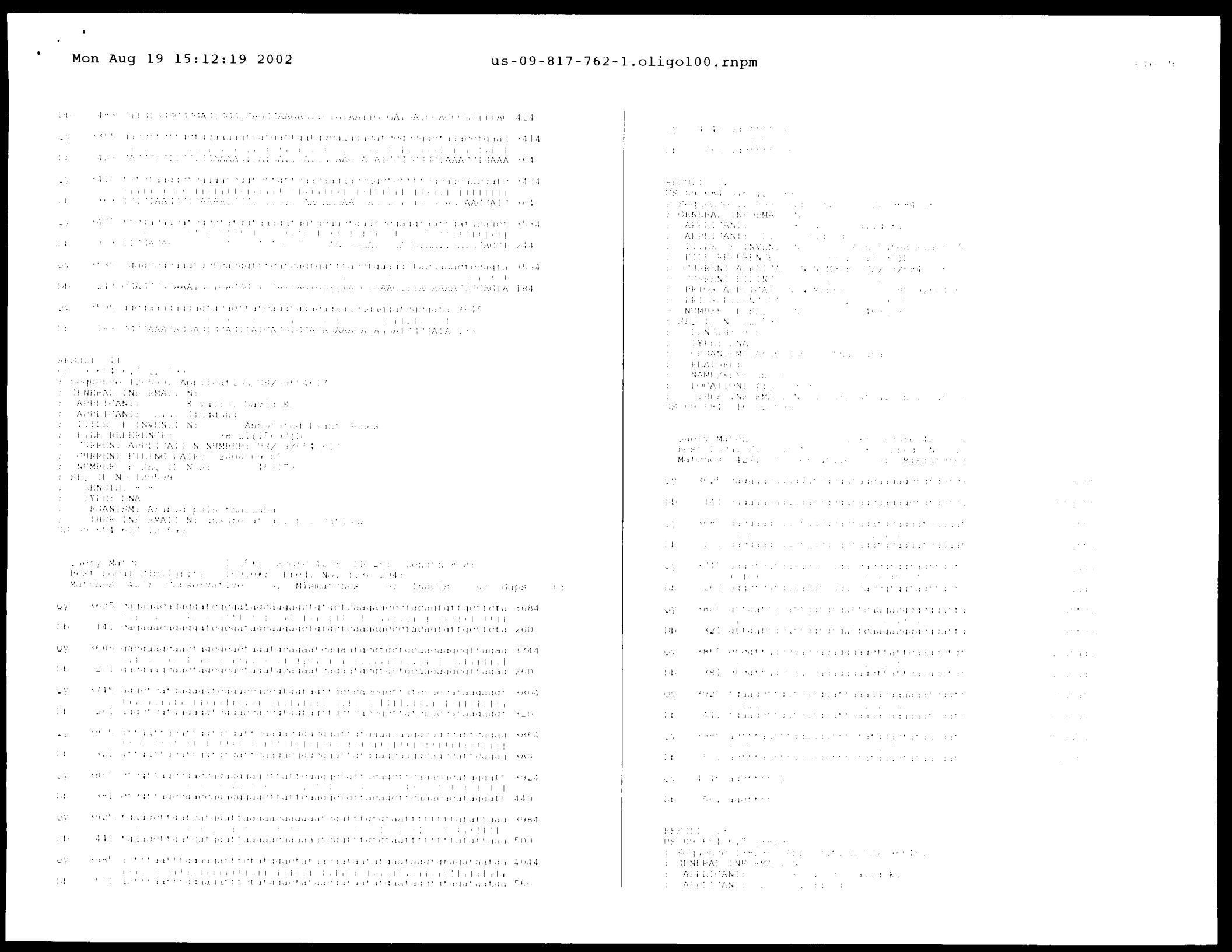


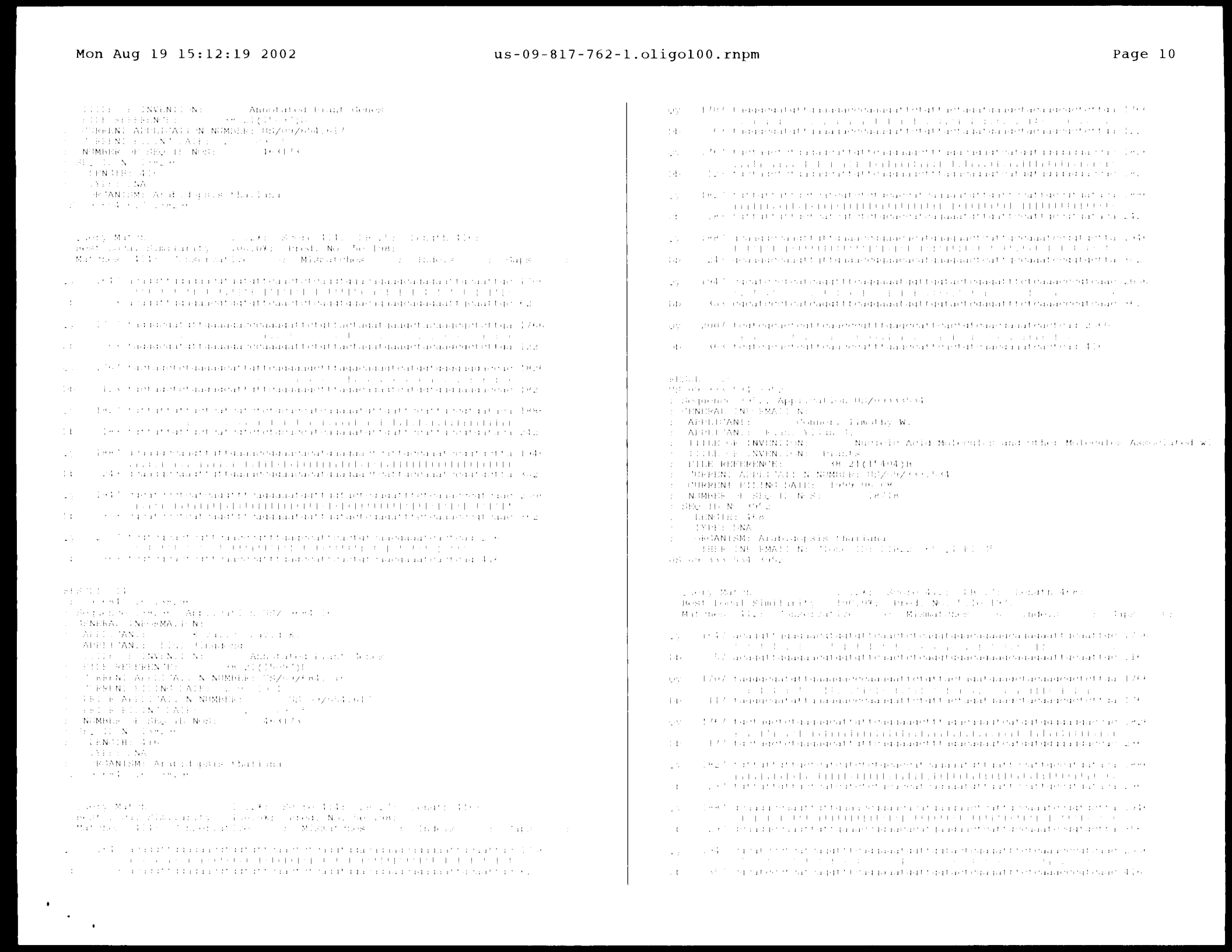


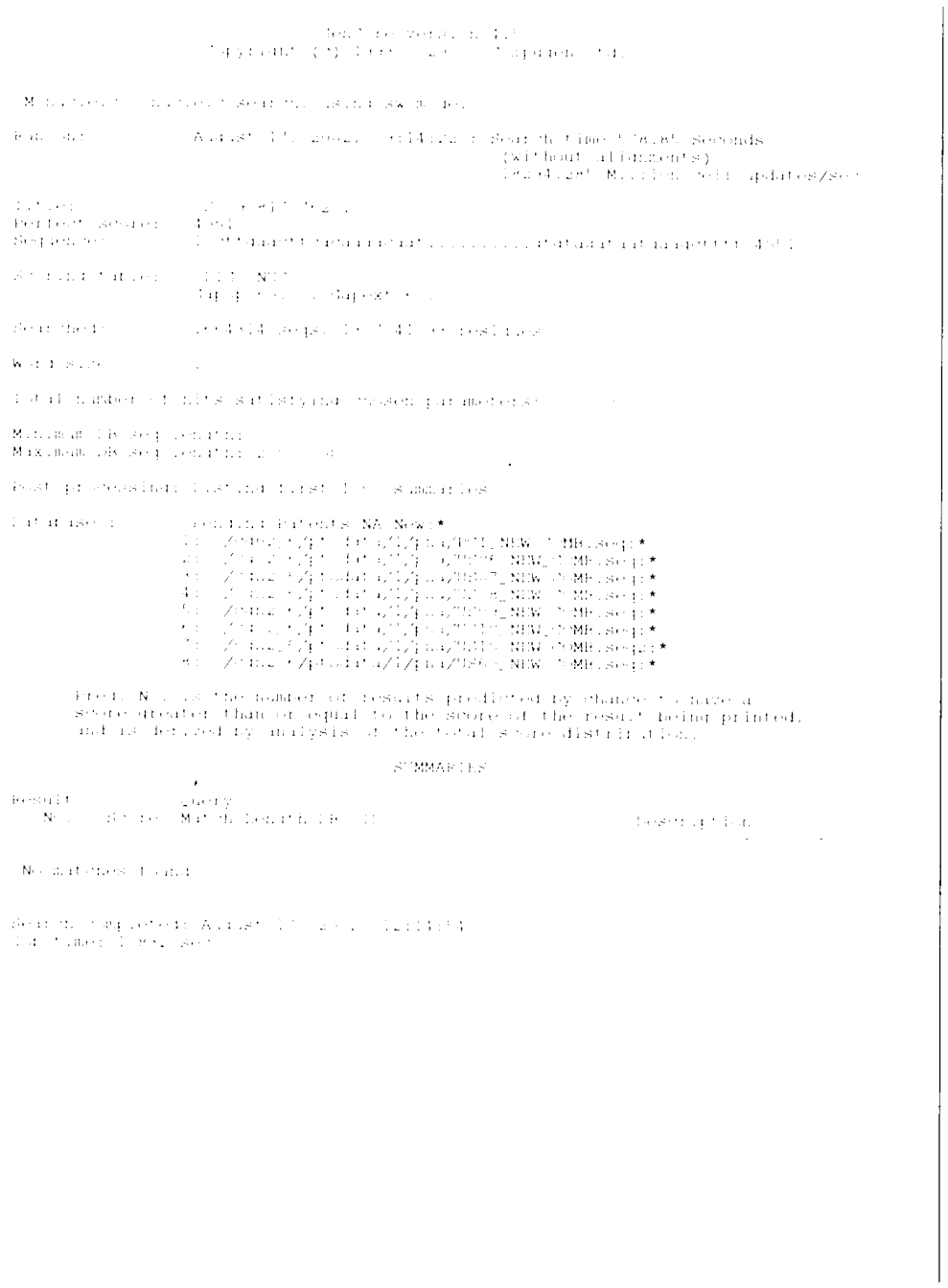


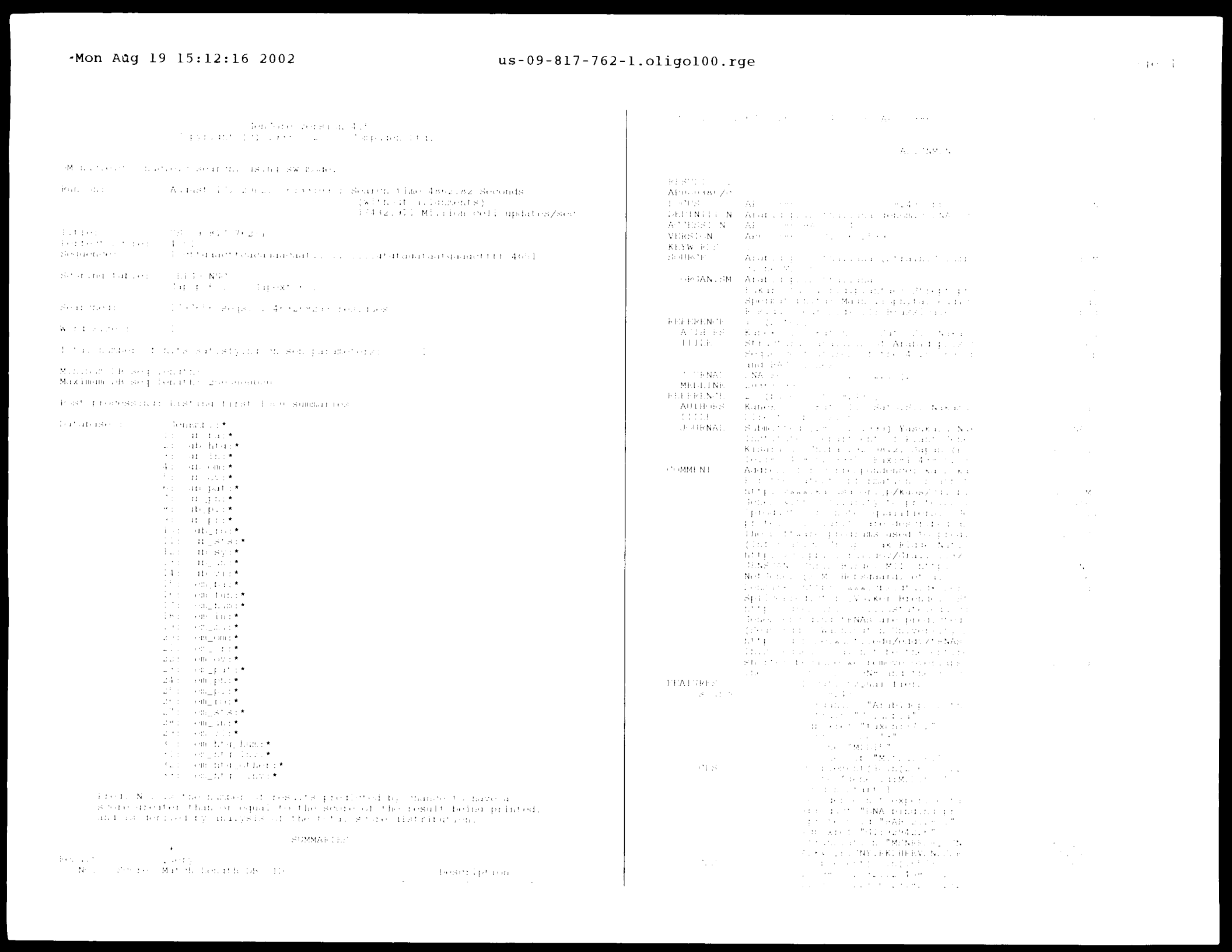


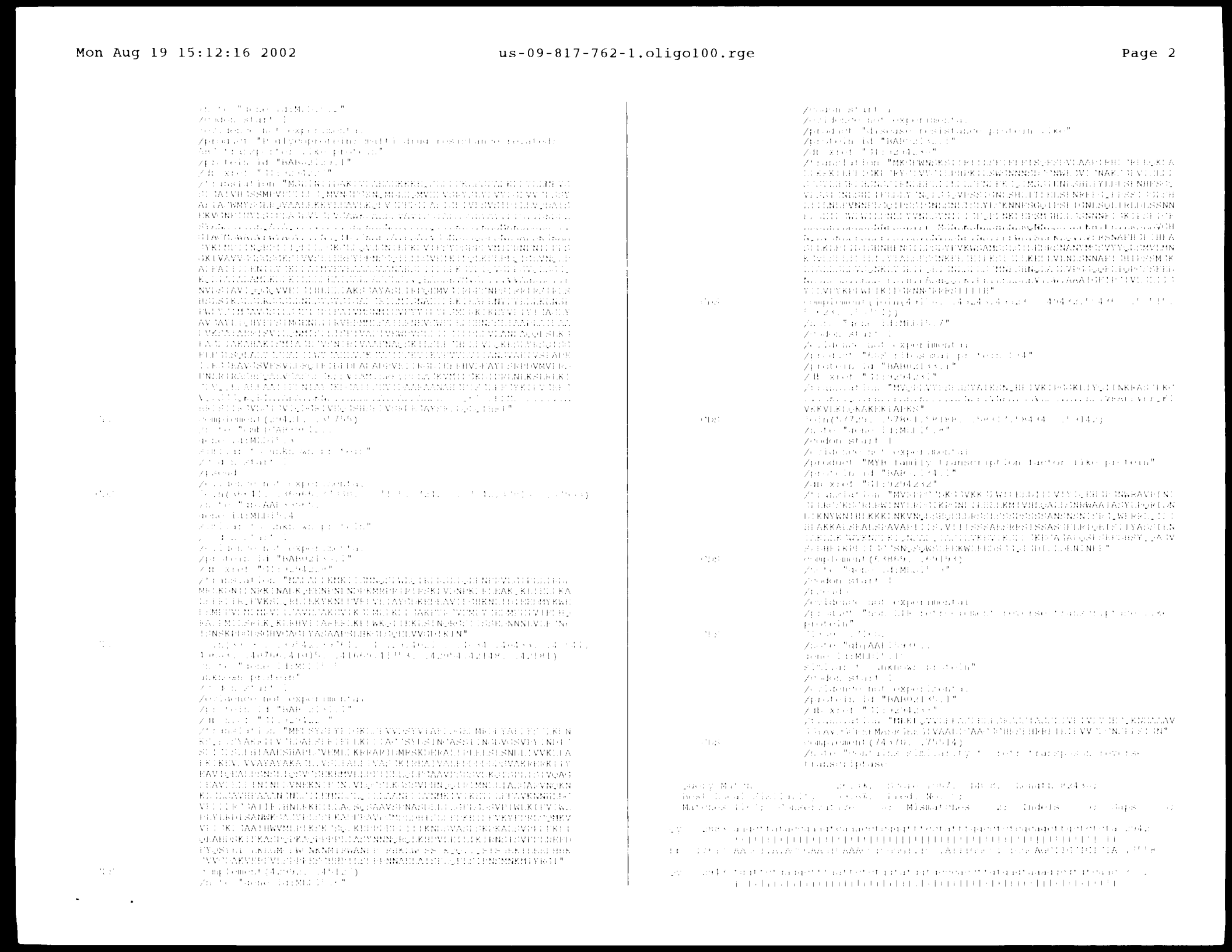


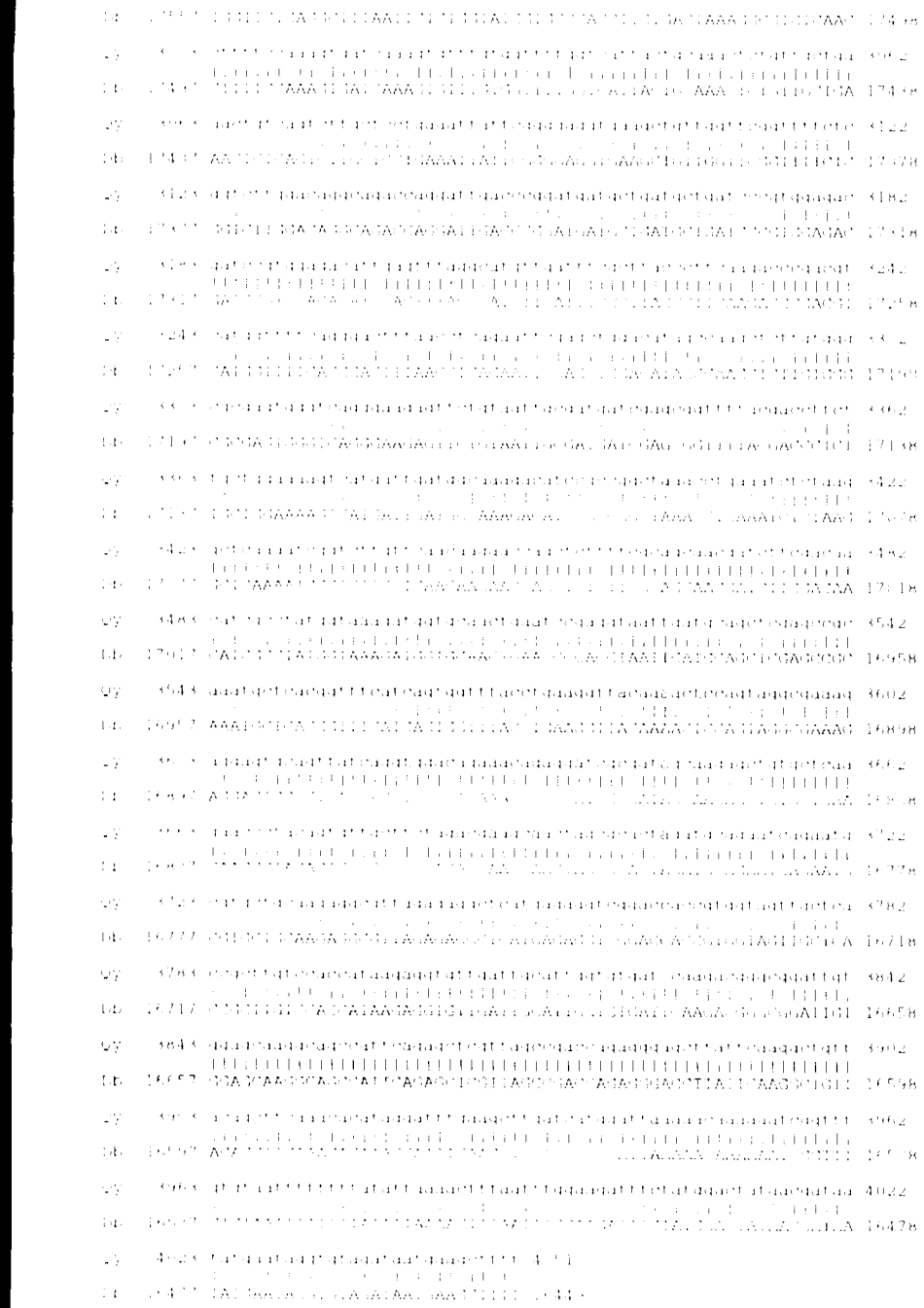


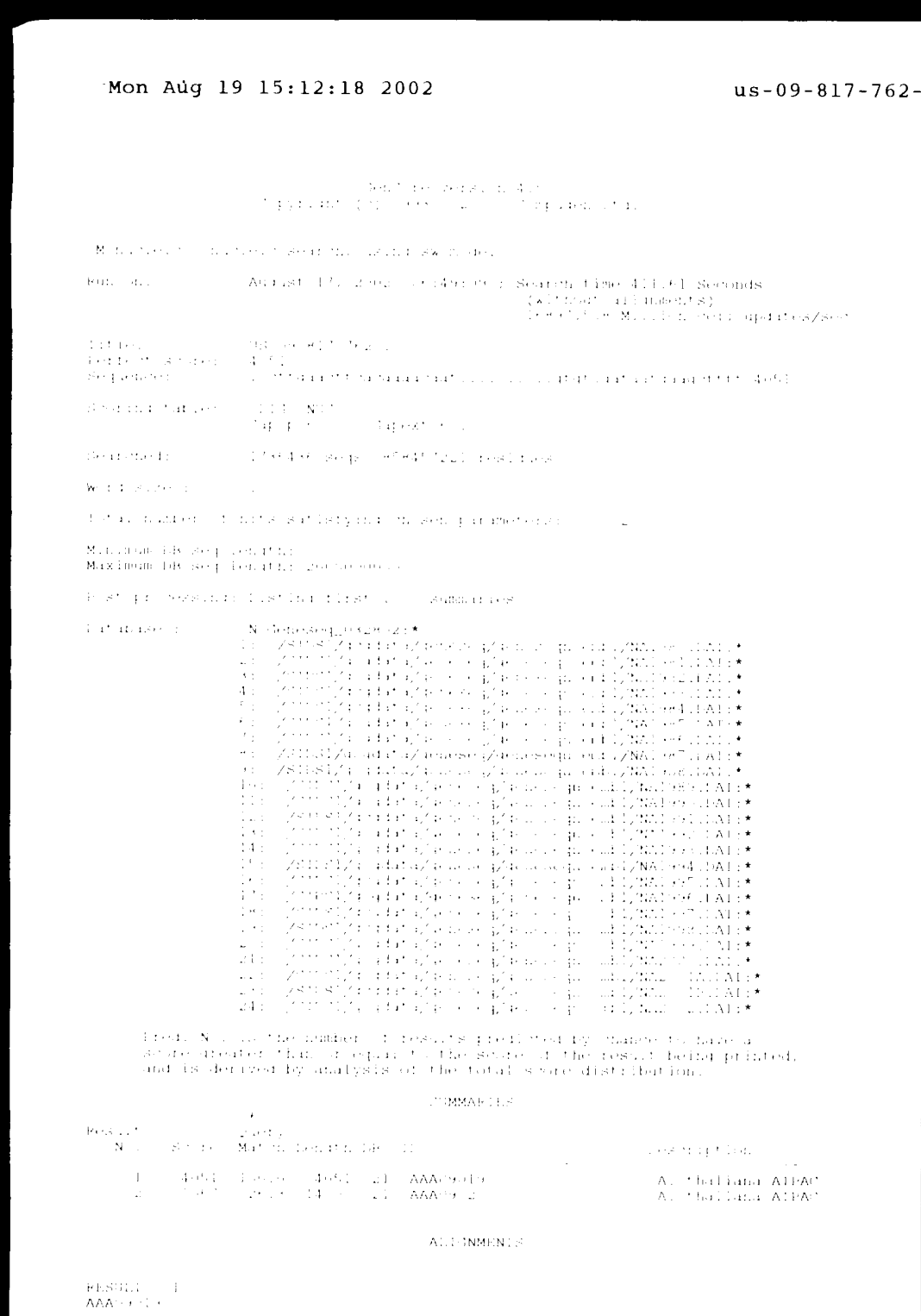


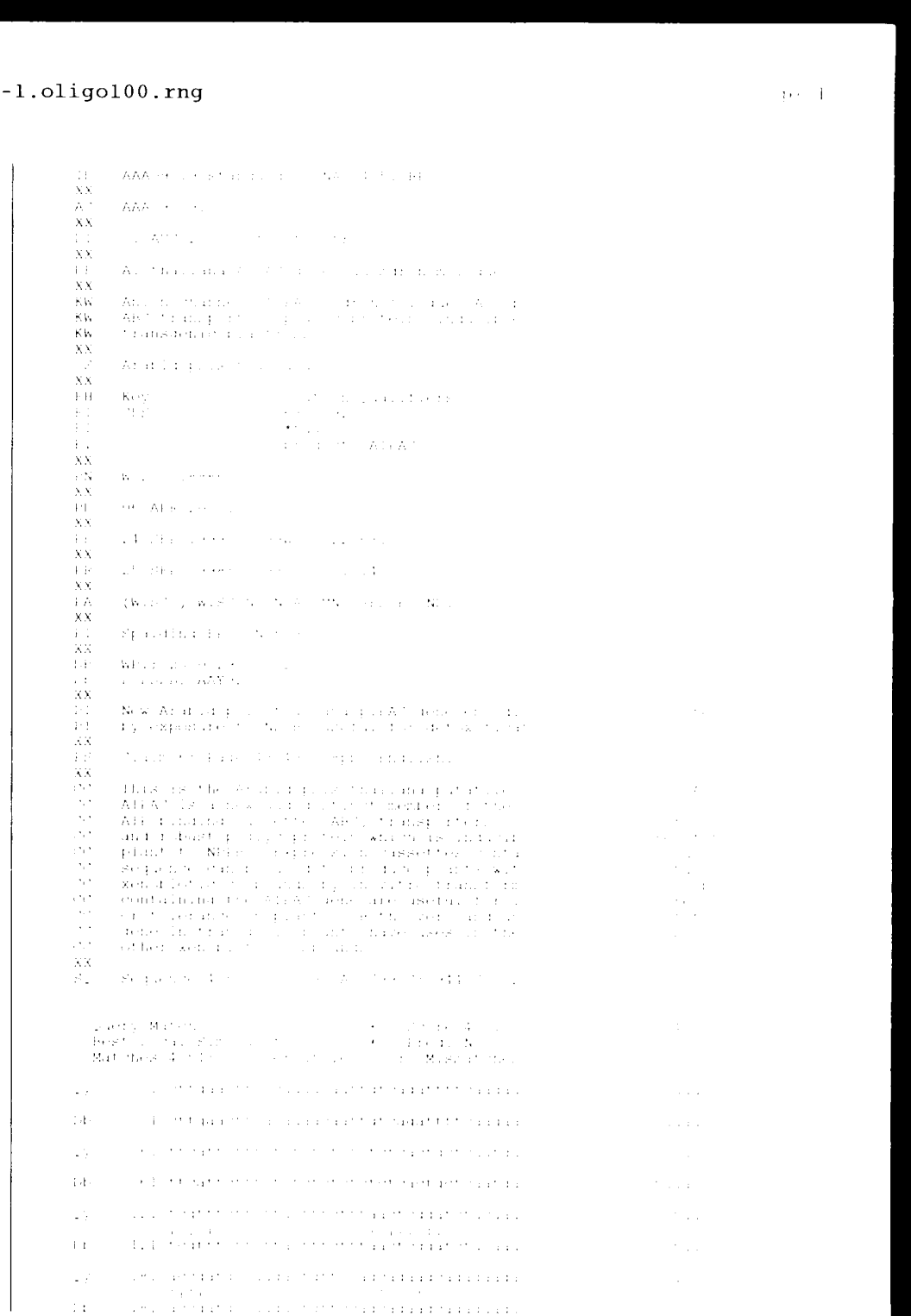


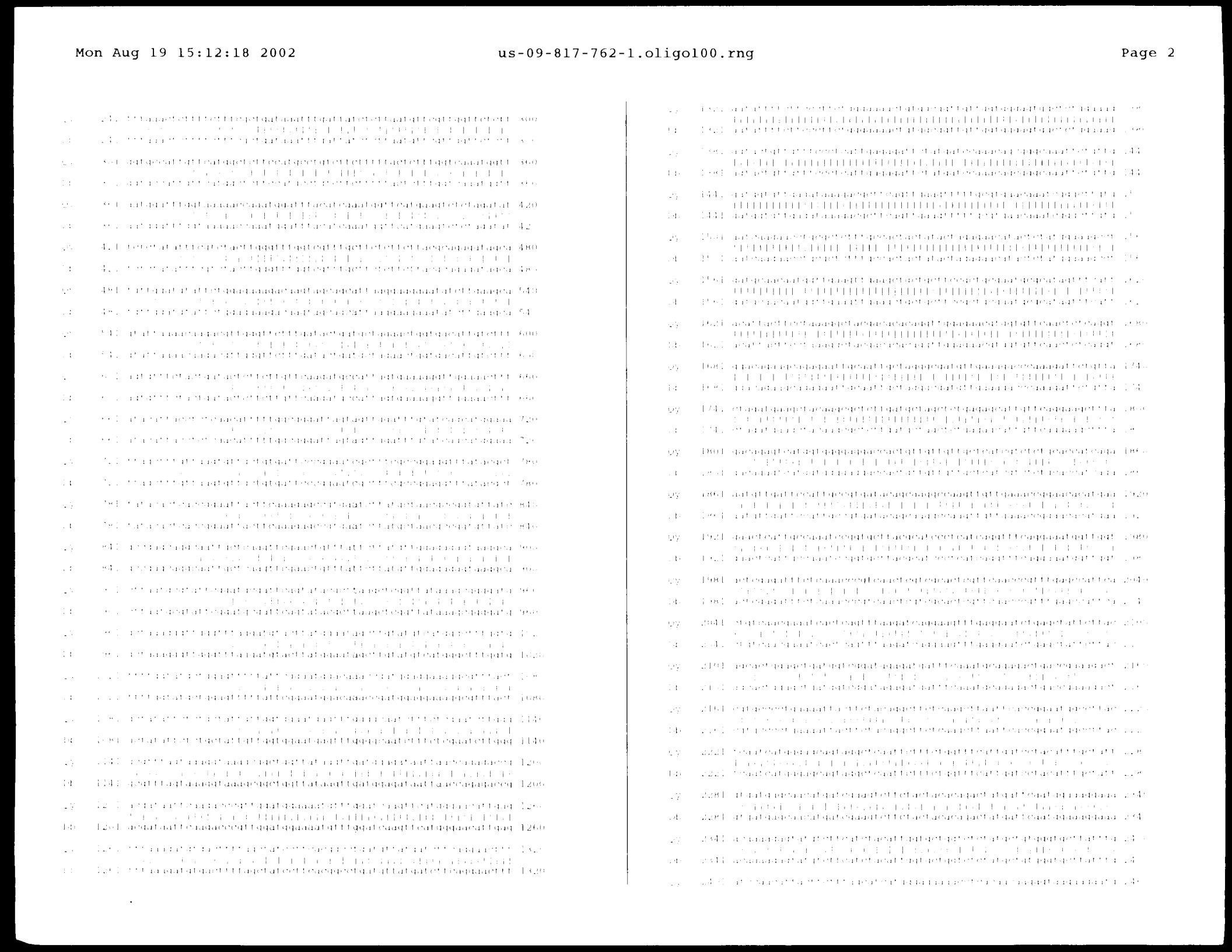


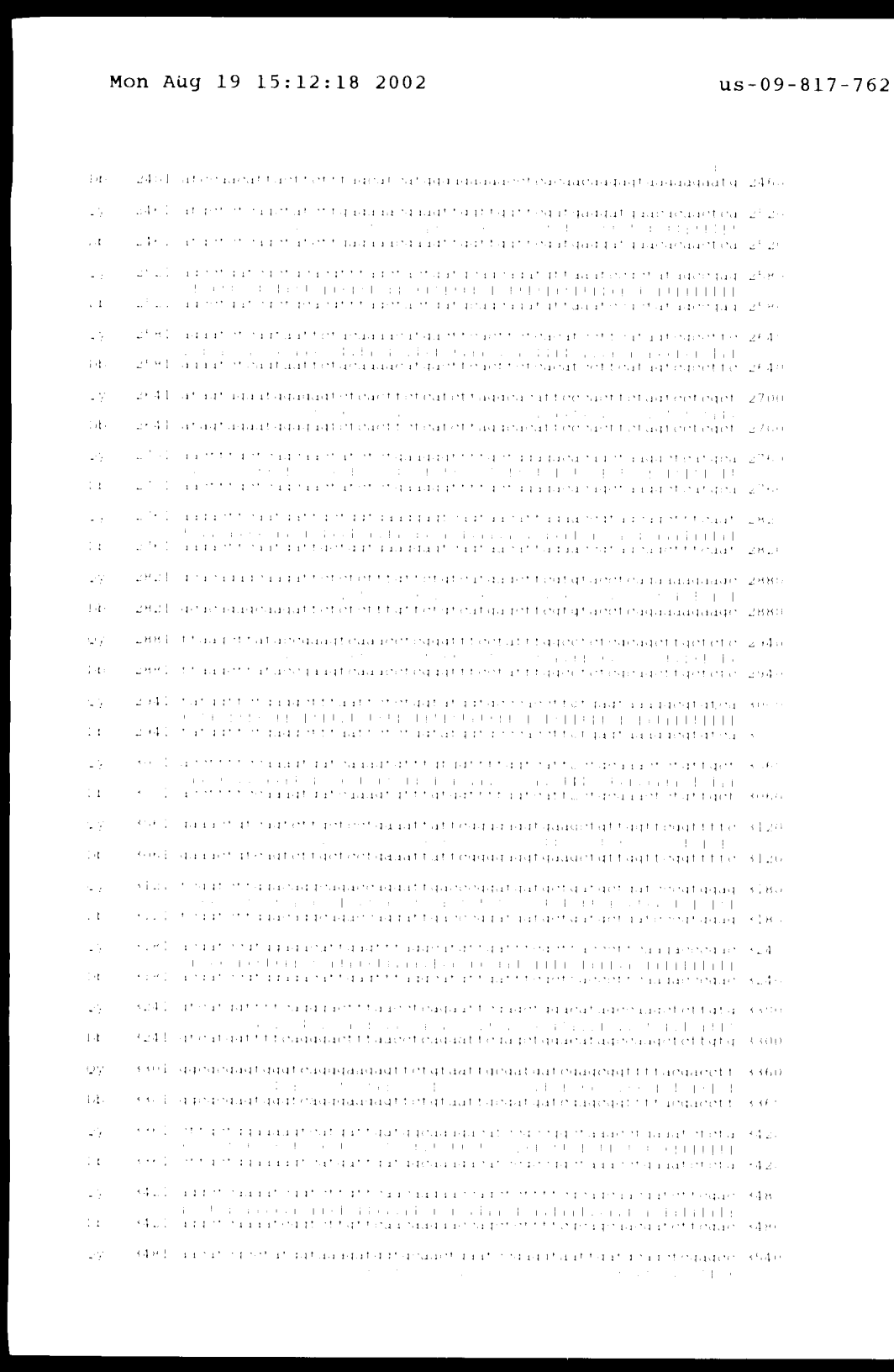


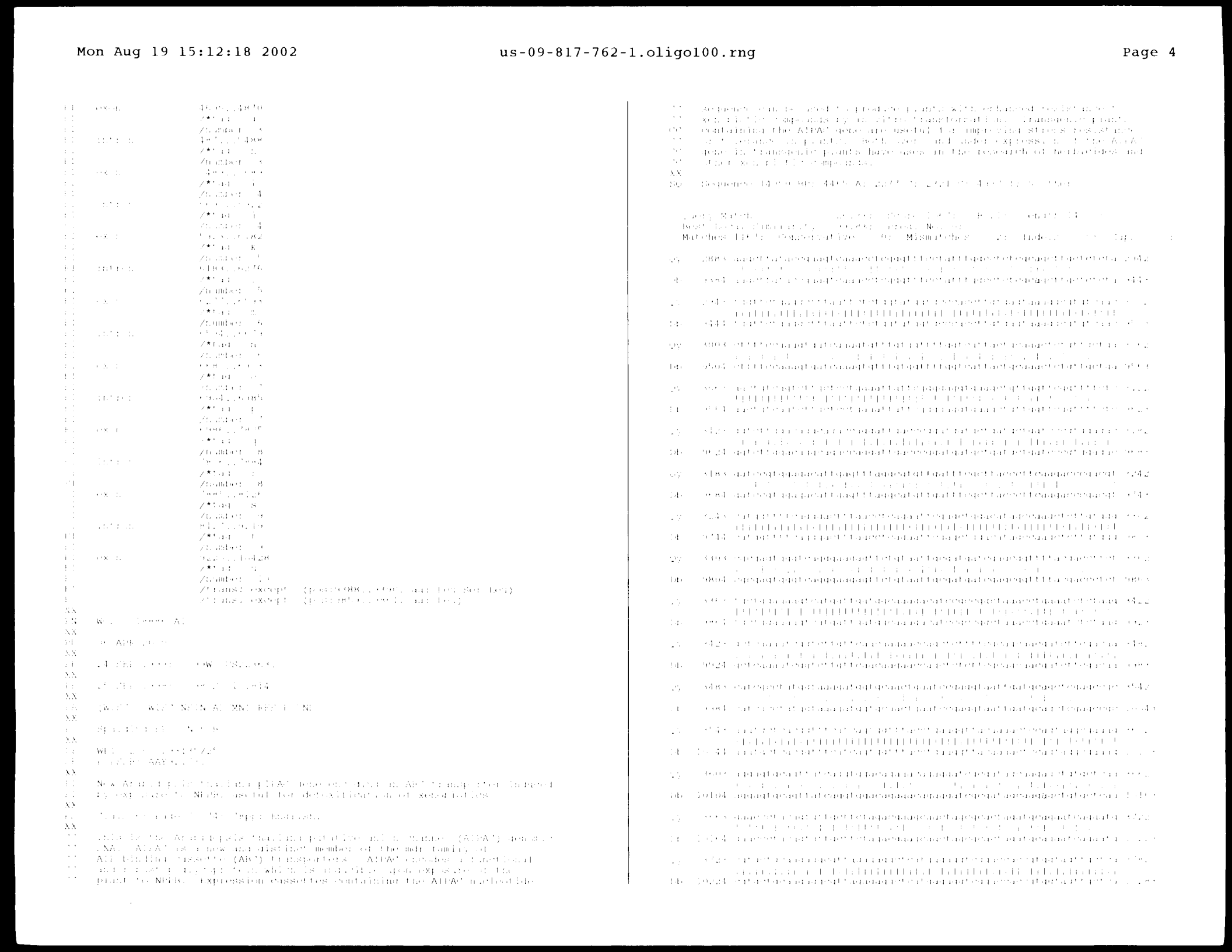


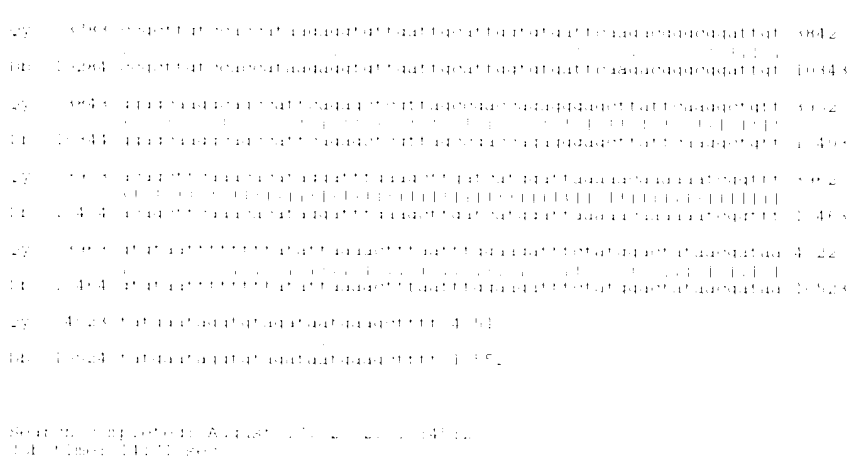


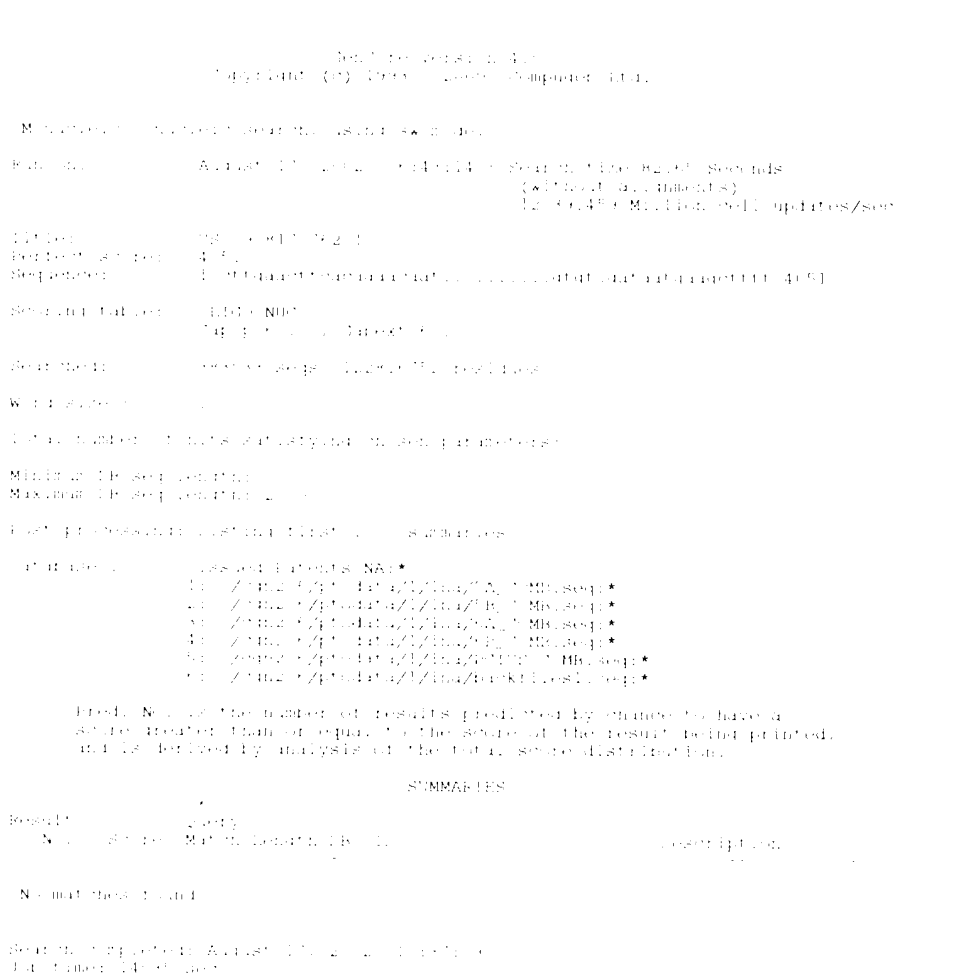












Copyright (c) [illegible] Compugen Ltd.

Run on: August [illegible] Search time 2042.17 Seconds
(without alignments)
[illegible] M[illegible] updates/sec

Title: [illegible] 817 762 1

Sequence: [illegible]tagaatgataatagaagctttt 4054

Word size: [illegible]

Total number of hits satisfying chosen parameters: [illegible]

Minimum DB seq length: [illegible]

Maximum DB seq length: [illegible]

Database: EST:*
[illegible] em_estba:*
[illegible] em_esthum:*
[illegible] em_estin:*
[illegible] em_estmu:*
[illegible] em_estov:*
[illegible] em_estpl:*
[illegible] em_estro:*
[illegible] em_htc:*
[illegible] gb_est1:*
[illegible] gb_est2:*
[illegible] gb_htc:*
[illegible] gb_gss:*
[illegible] em_gss_hum:*
[illegible] em_gss_inv:*
[illegible] em_gss_pln:*
[illegible] em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV547446	AV547446 AV547446
2	[illegible]	[illegible]	[illegible]	[illegible]	AV541481	AV541481 AV541481
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV543928	AV543928 AV543928
4	[illegible]	[illegible]	[illegible]	[illegible]	AA605450	AA605450 30486 Lam
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV552386	AV552386 AV552386
[illegible]	[illegible]	[illegible]	475	[illegible]	AV547920	AV547920 AV547920
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV553581	AV553581 AV553581
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AI994333	AI994333 701502878
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV548435	AV548435 AV548435
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	N95860	N95860 22002 Lambd
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	BE528318	BE528318 M7K21SIM
12	[illegible]	[illegible]	[illegible]	[illegible]	BE529177	BE529177 M7[illegible]SIM
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AA394594	AA394594 [illegible] Lam
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	BE533661	BE533661 M78AG[illegible]SIM
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	AV546402	AV546402 AV546402

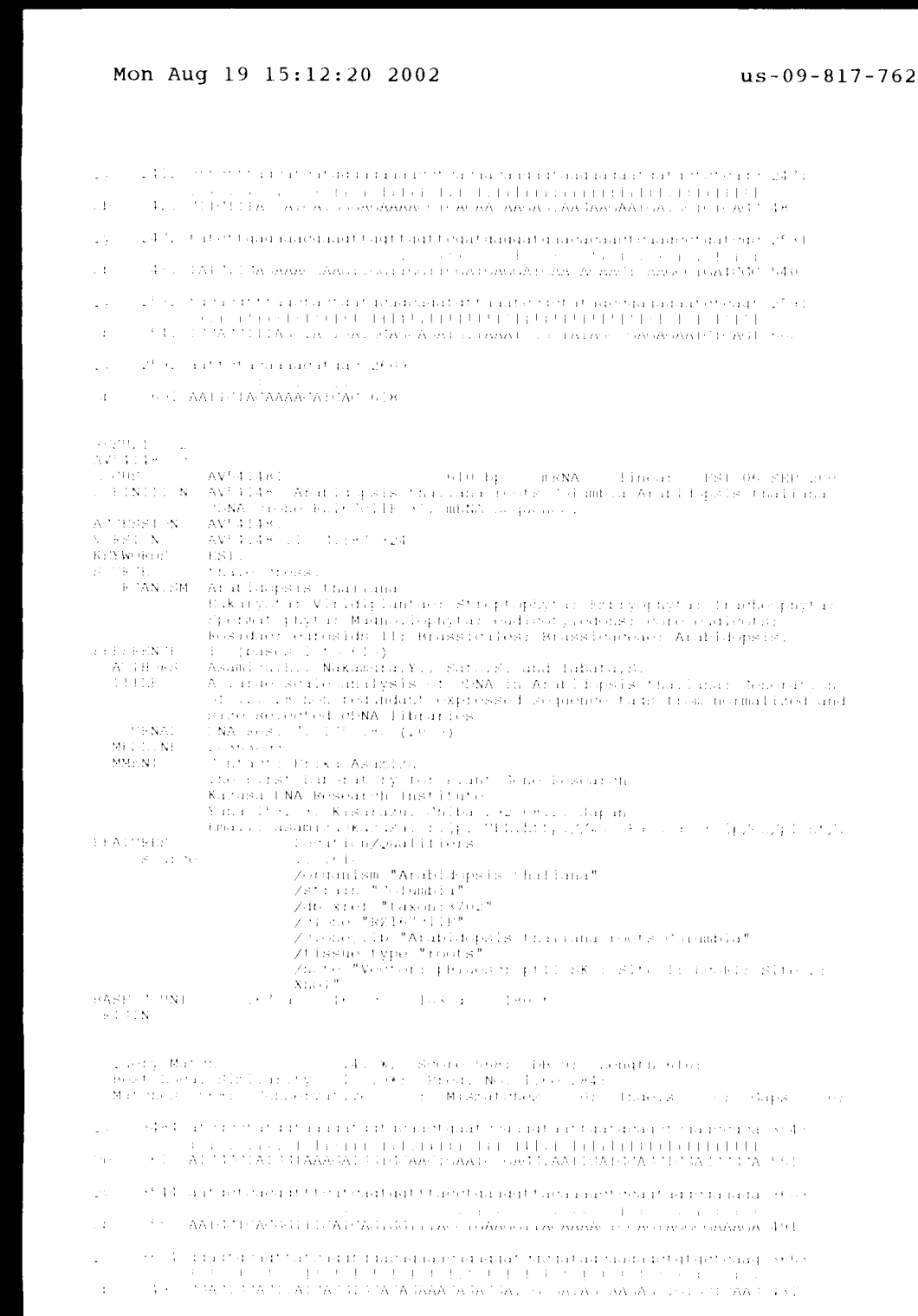
ALIGNMENTS

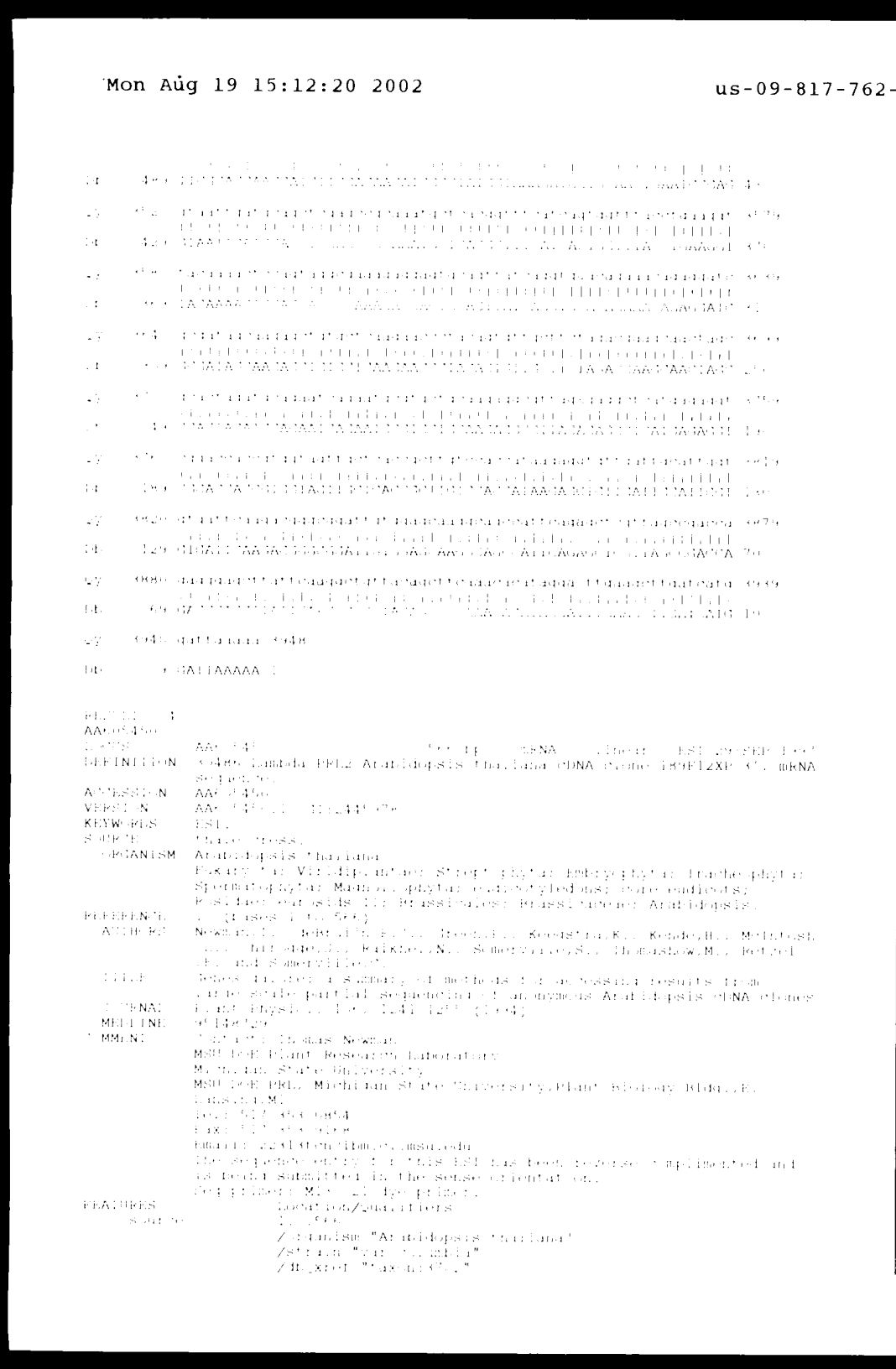
RESULT [illegible]
AV547446
LOCUS AV5[illegible]
DEFINITION AV5[illegible]
[illegible] mRNA [illegible]
ACCESSION AV5[illegible]
VERSION AV5[illegible]
KEYWORDS EST.
SOURCE [illegible]
ORGANISM Arab[illegible]
REFERENCE [illegible]
AUTHORS [illegible]
TITLE [illegible]
JOURNAL [illegible]
MEDLINE [illegible]
COMMENT [illegible]
FEATURES [illegible]
source [illegible]
BASE COUNT [illegible]
ORIGIN

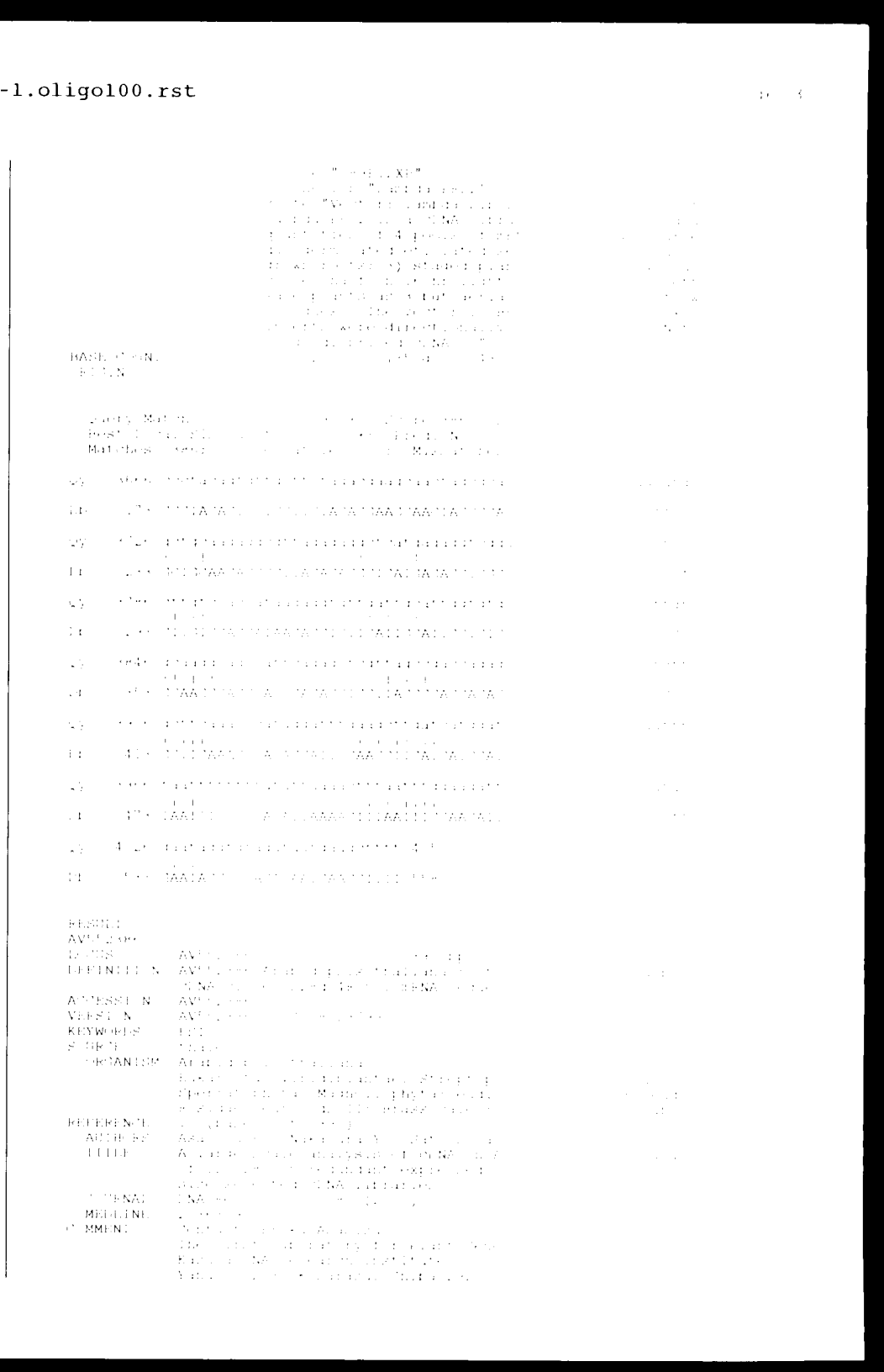
Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible]

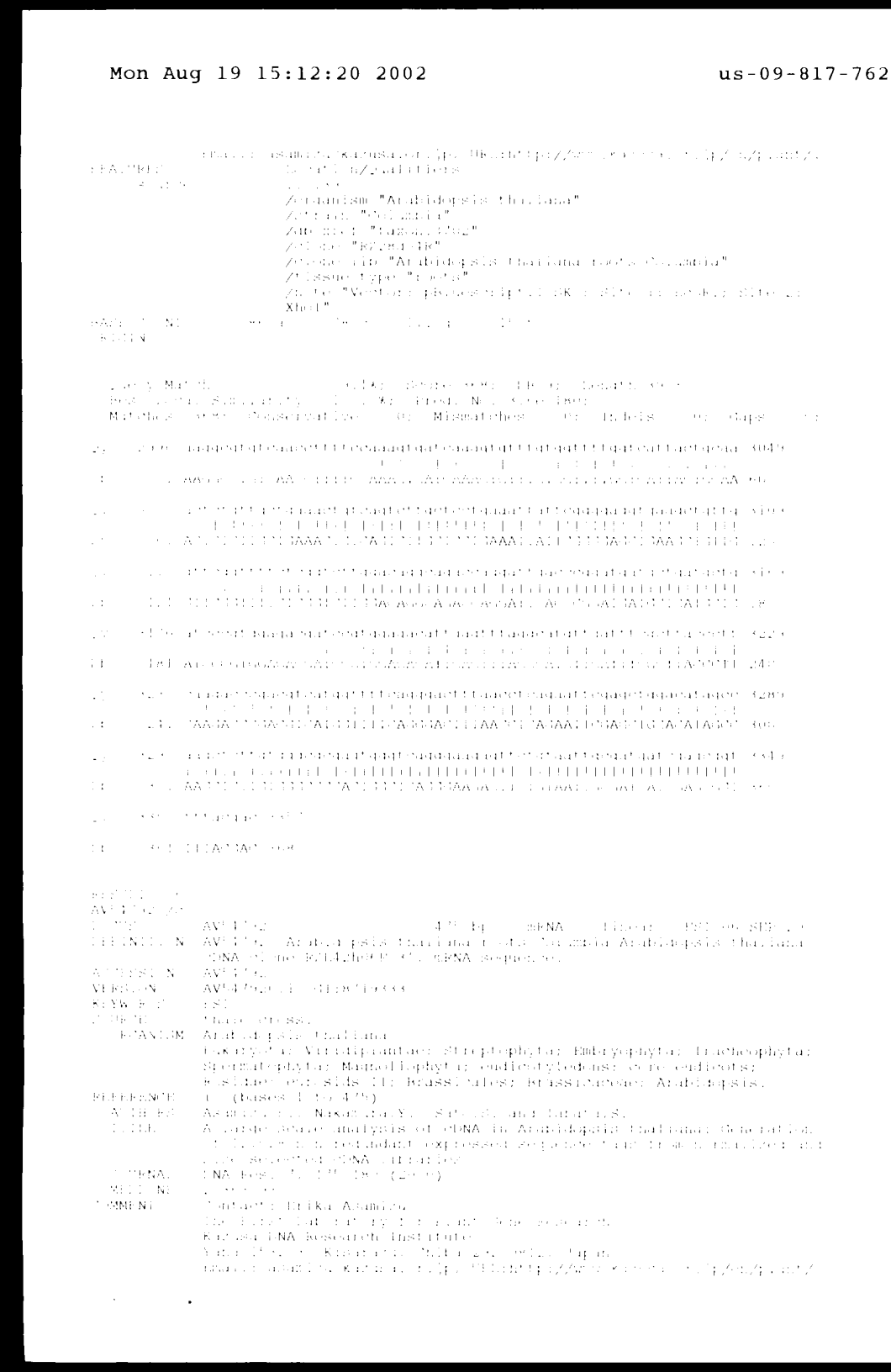
Qy [illegible]

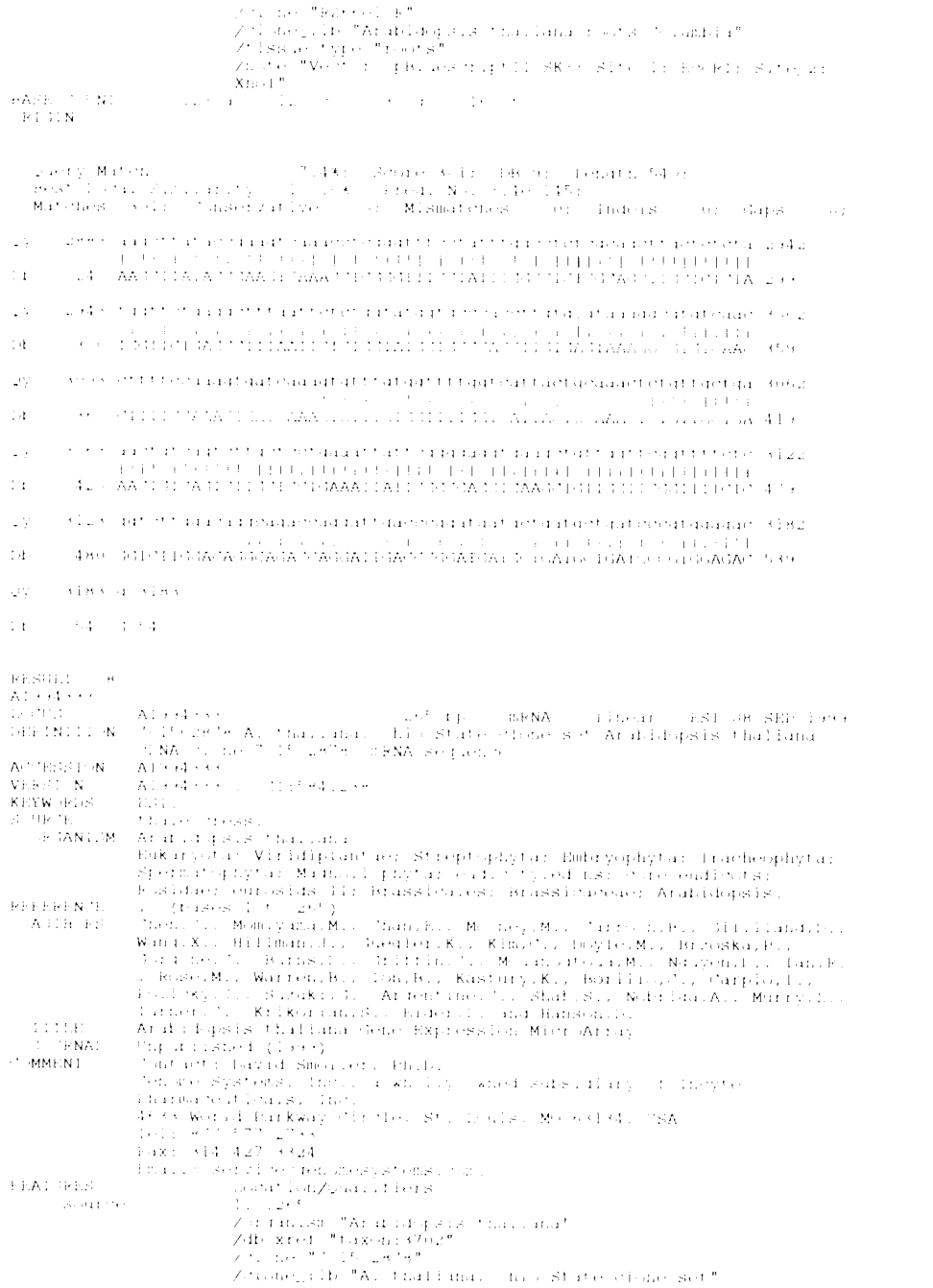
Db [illegible]

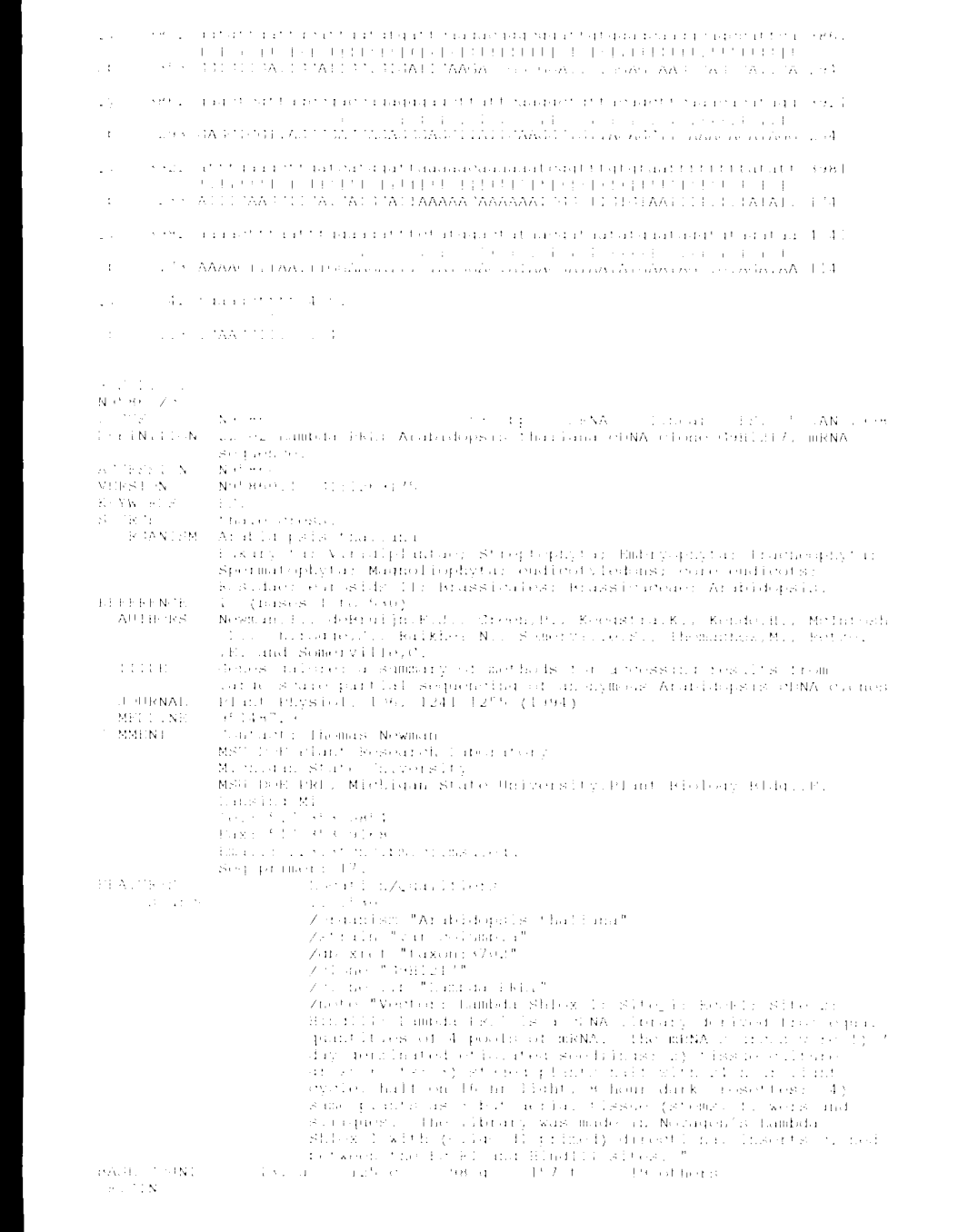


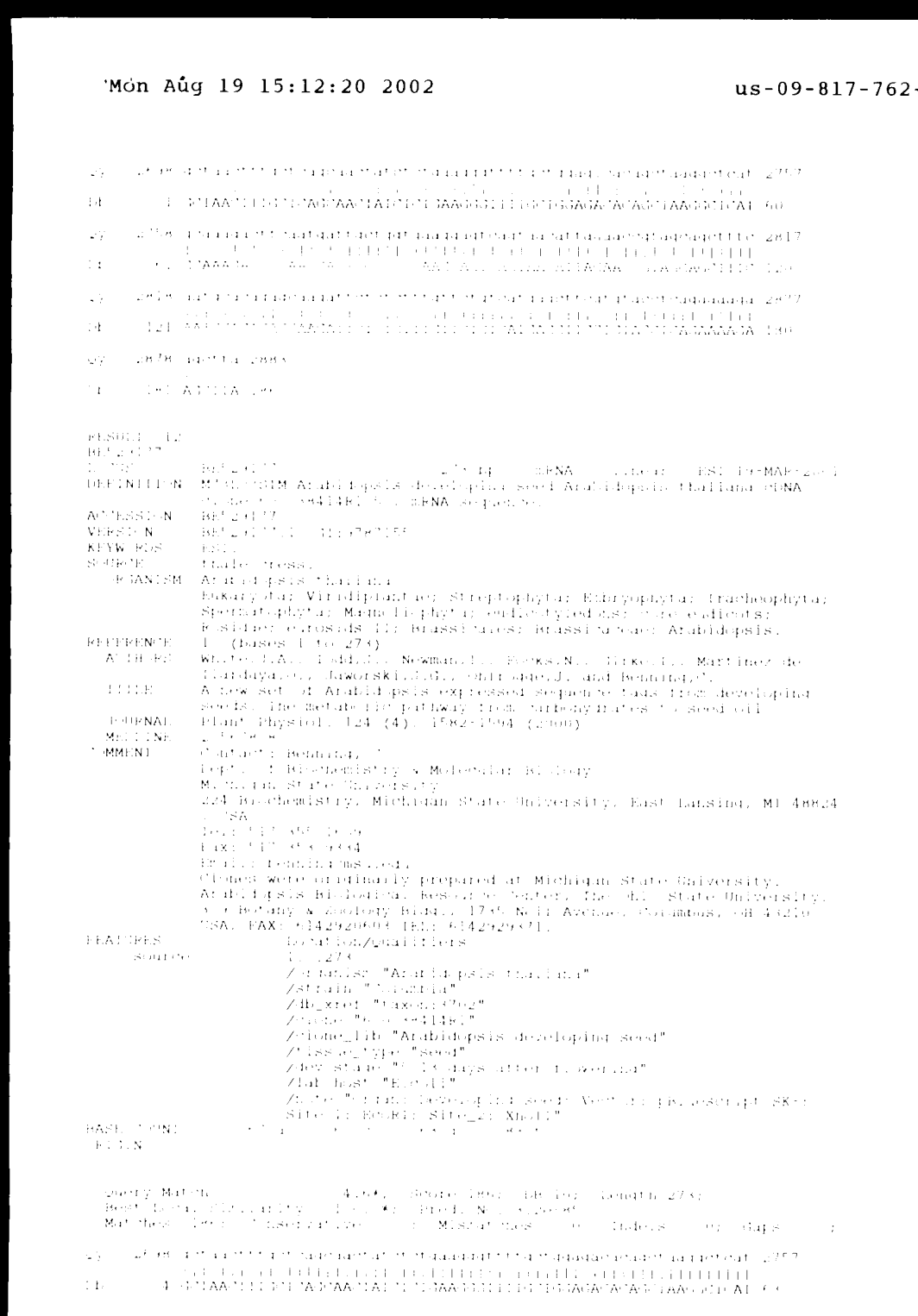


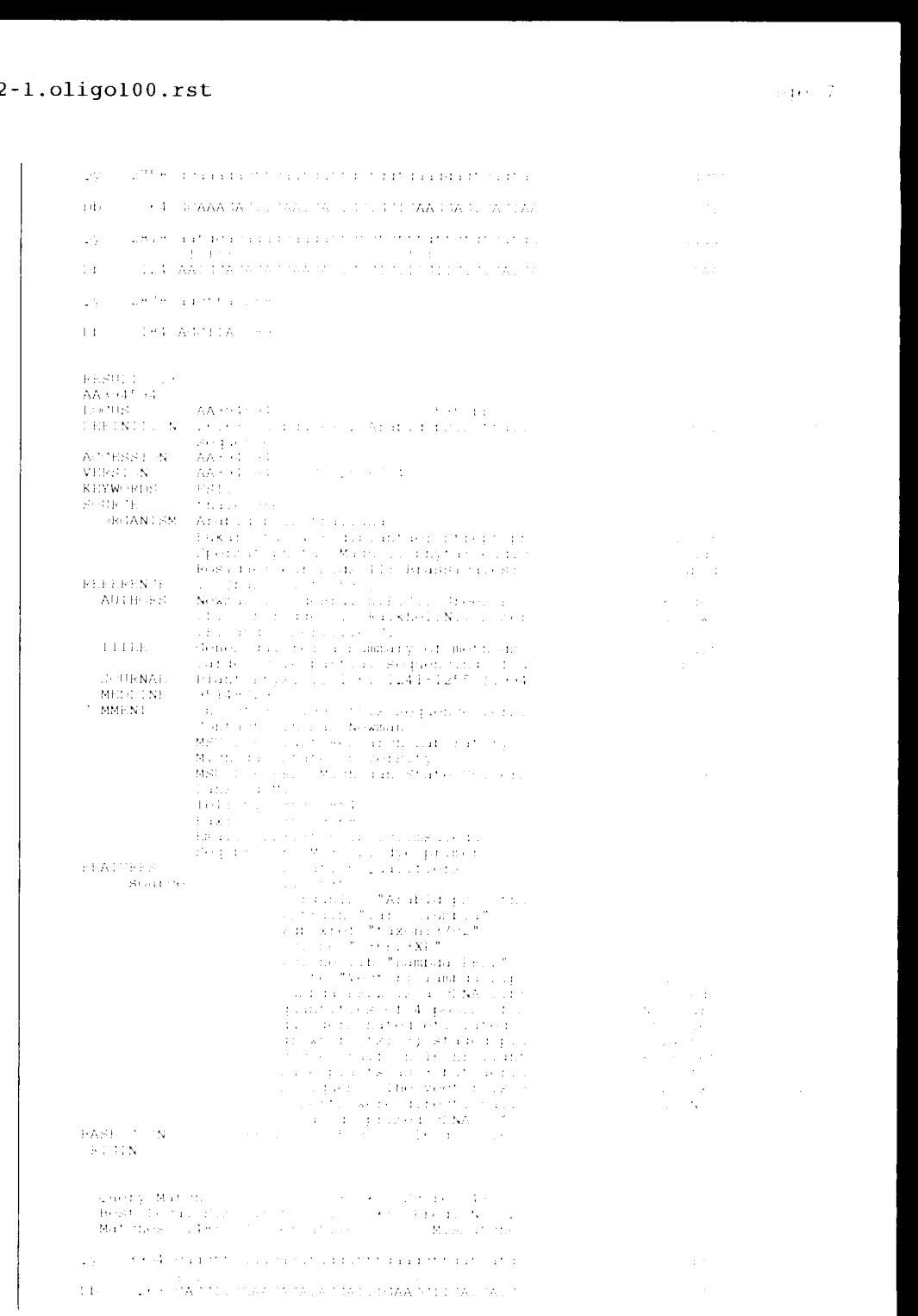


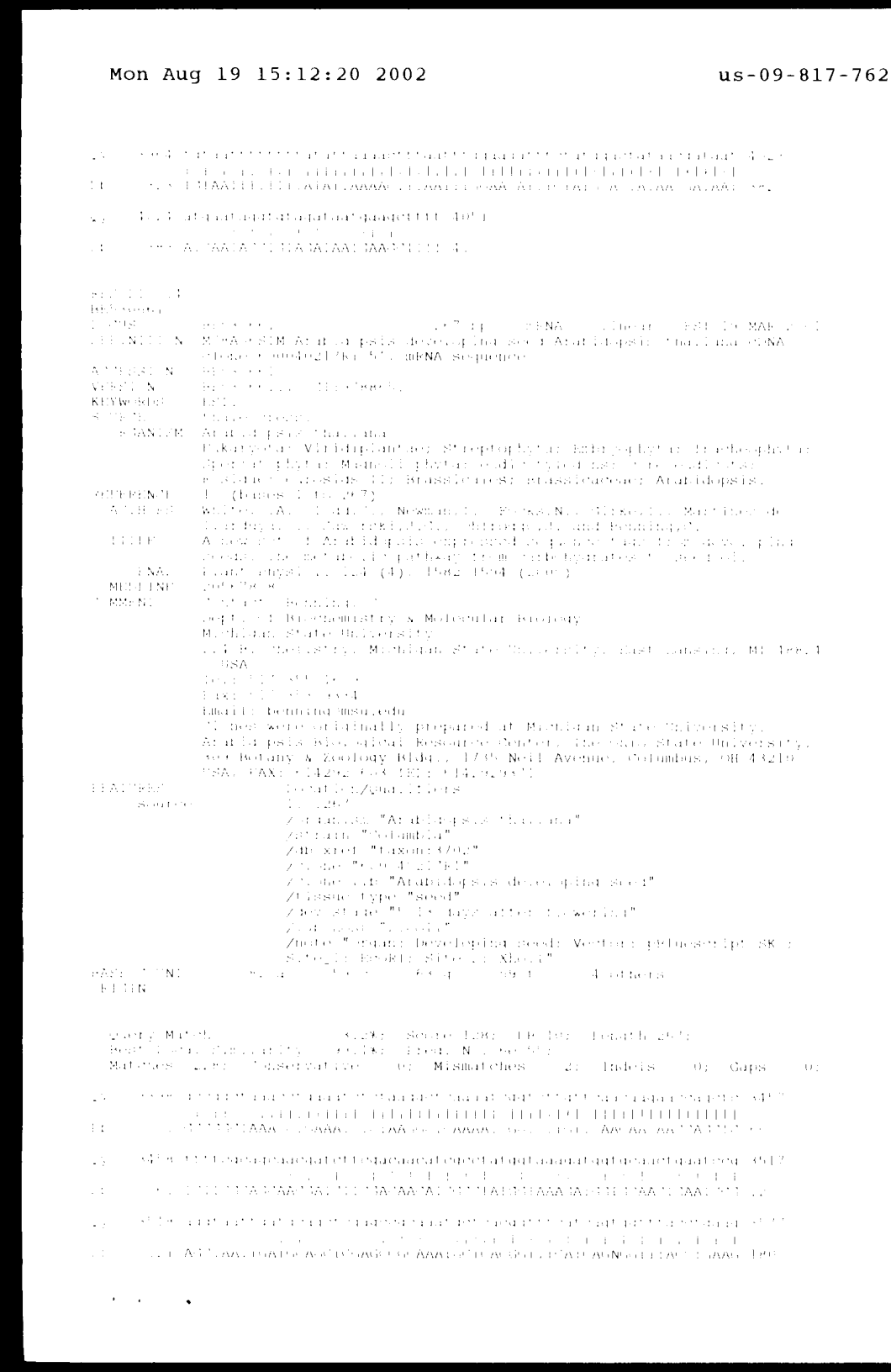


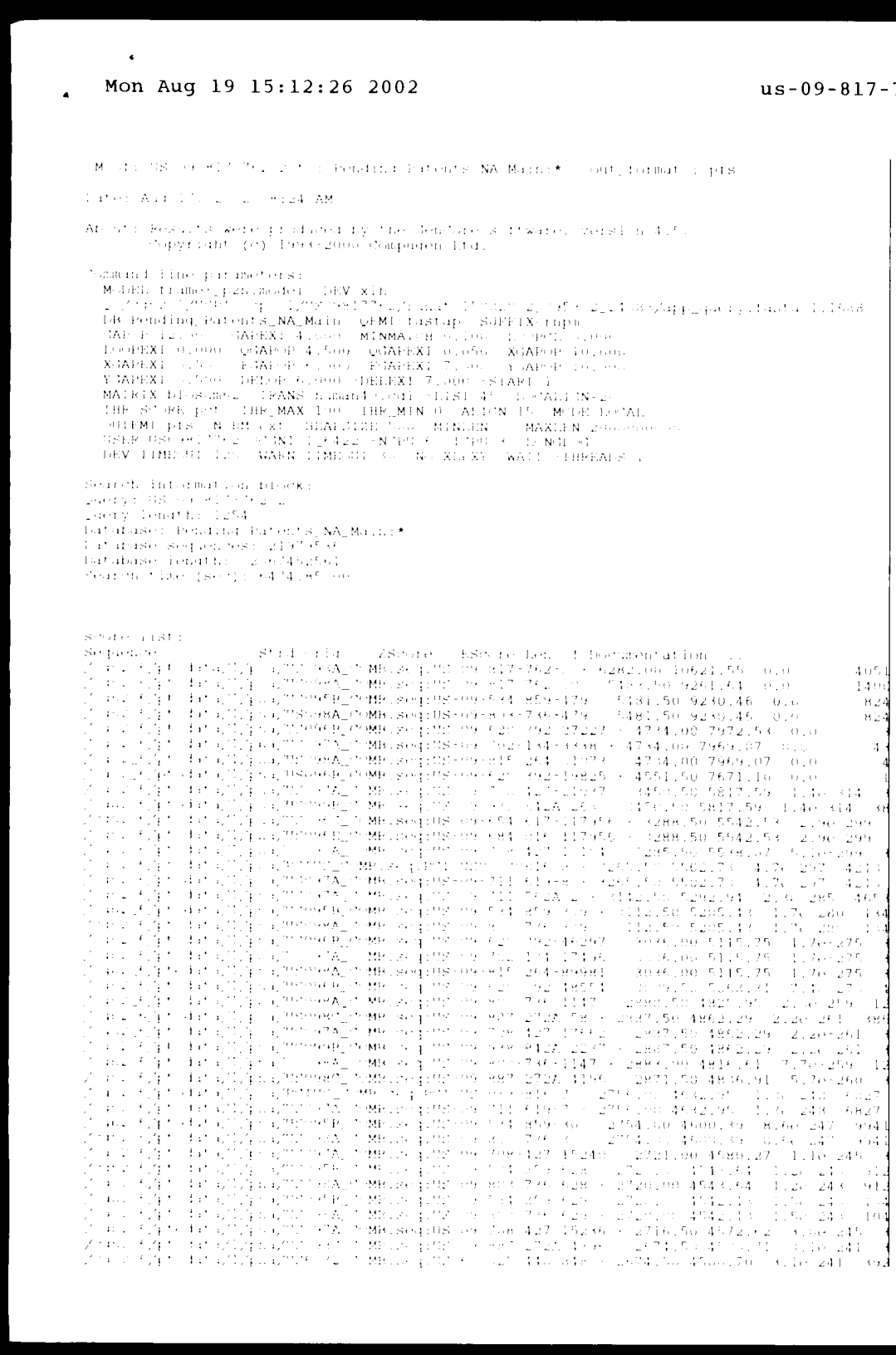


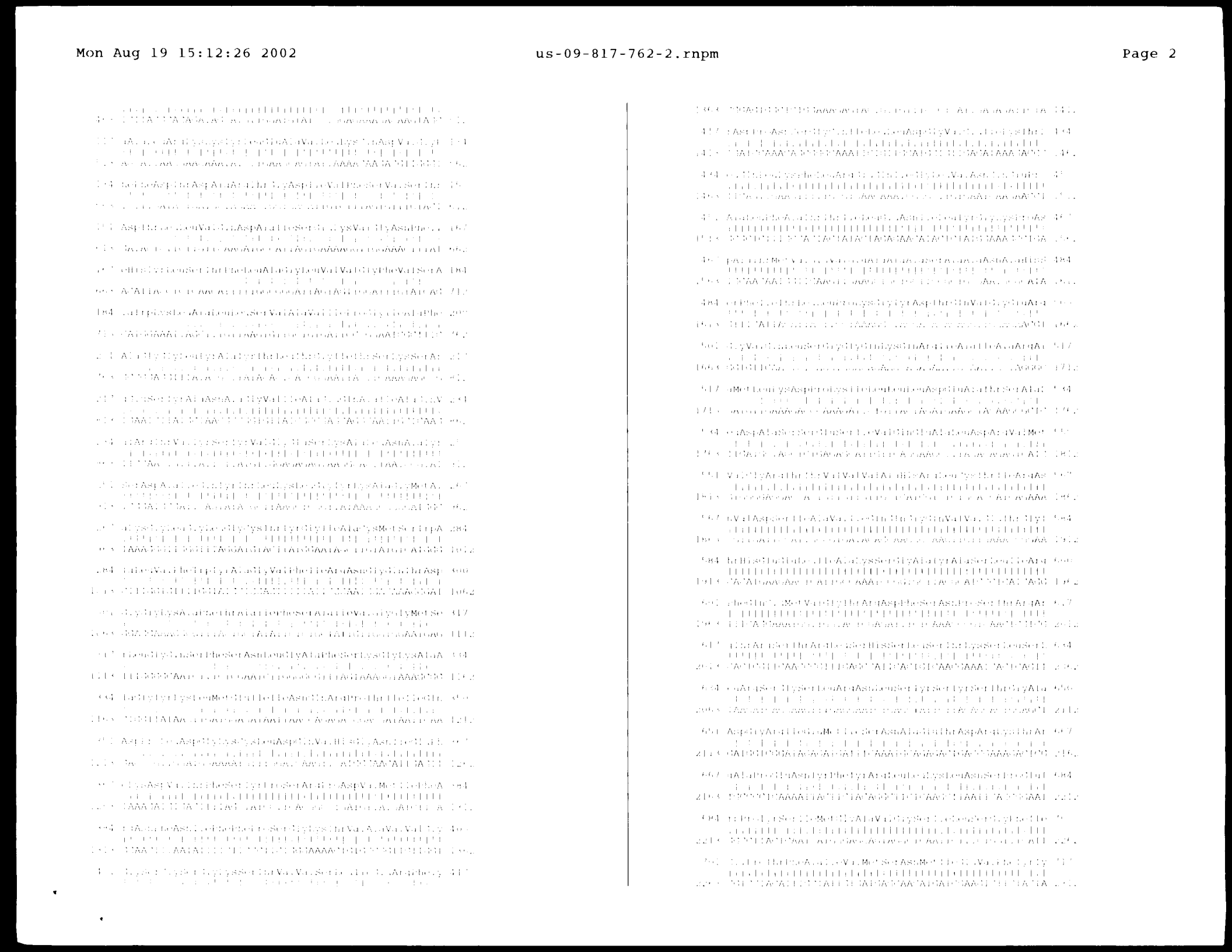


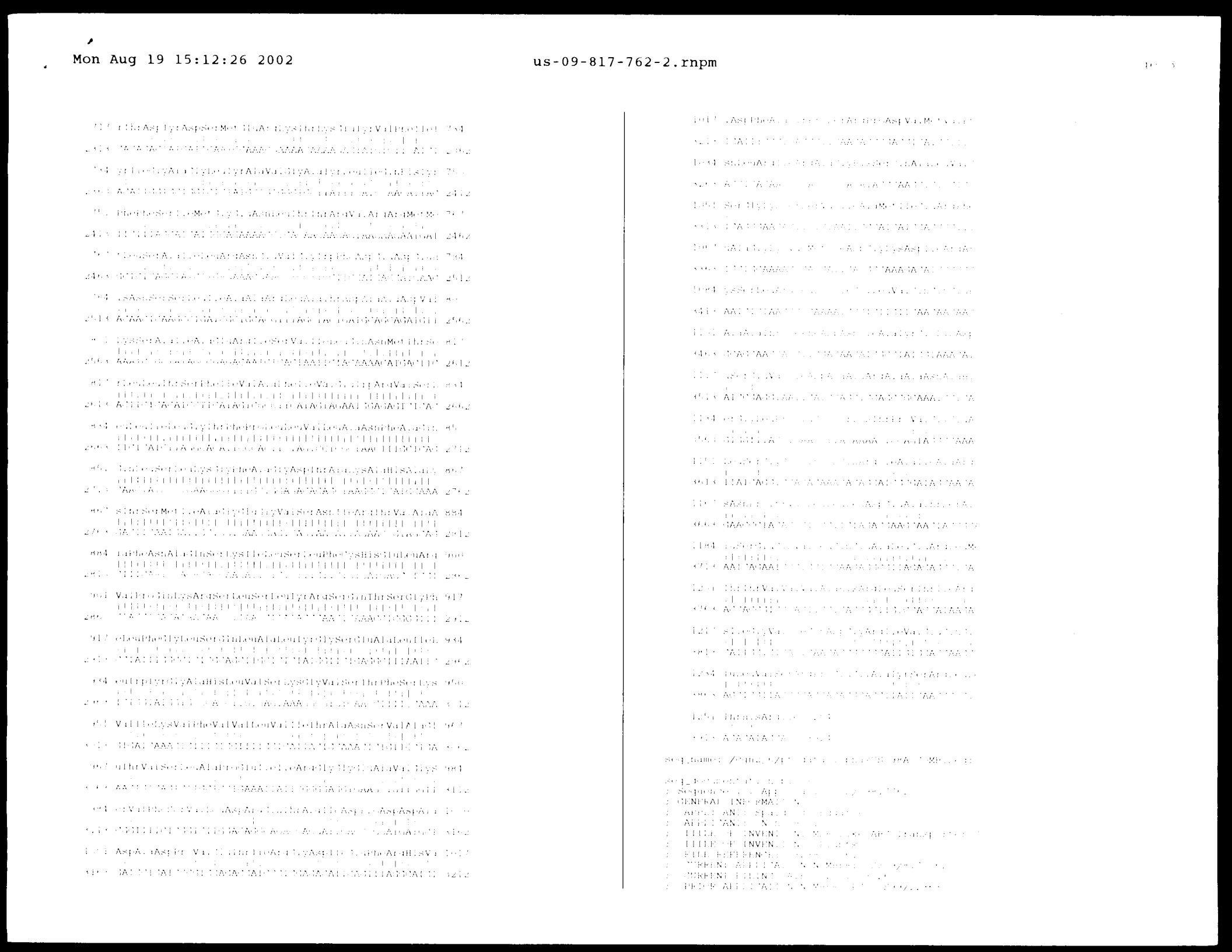


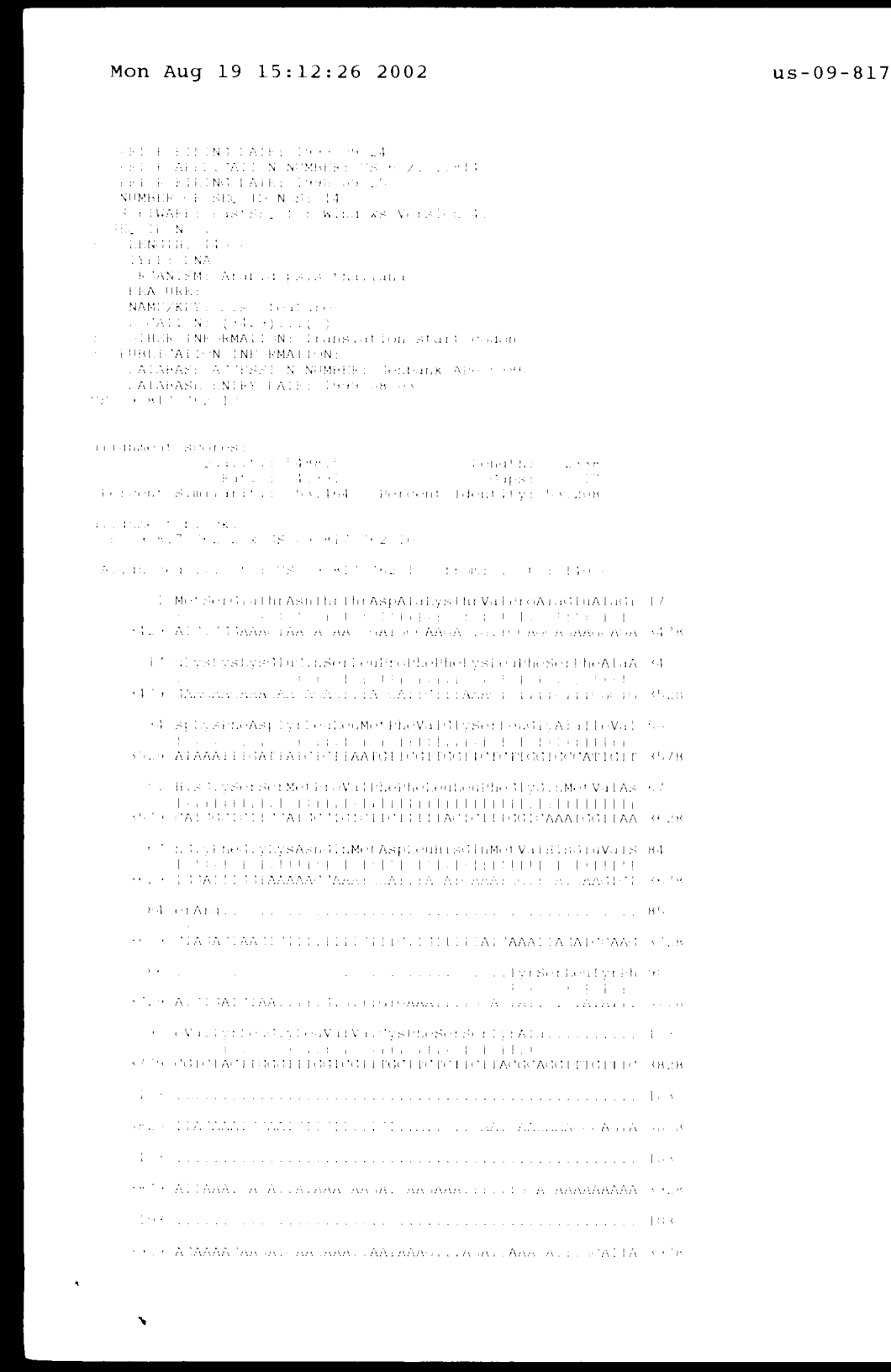


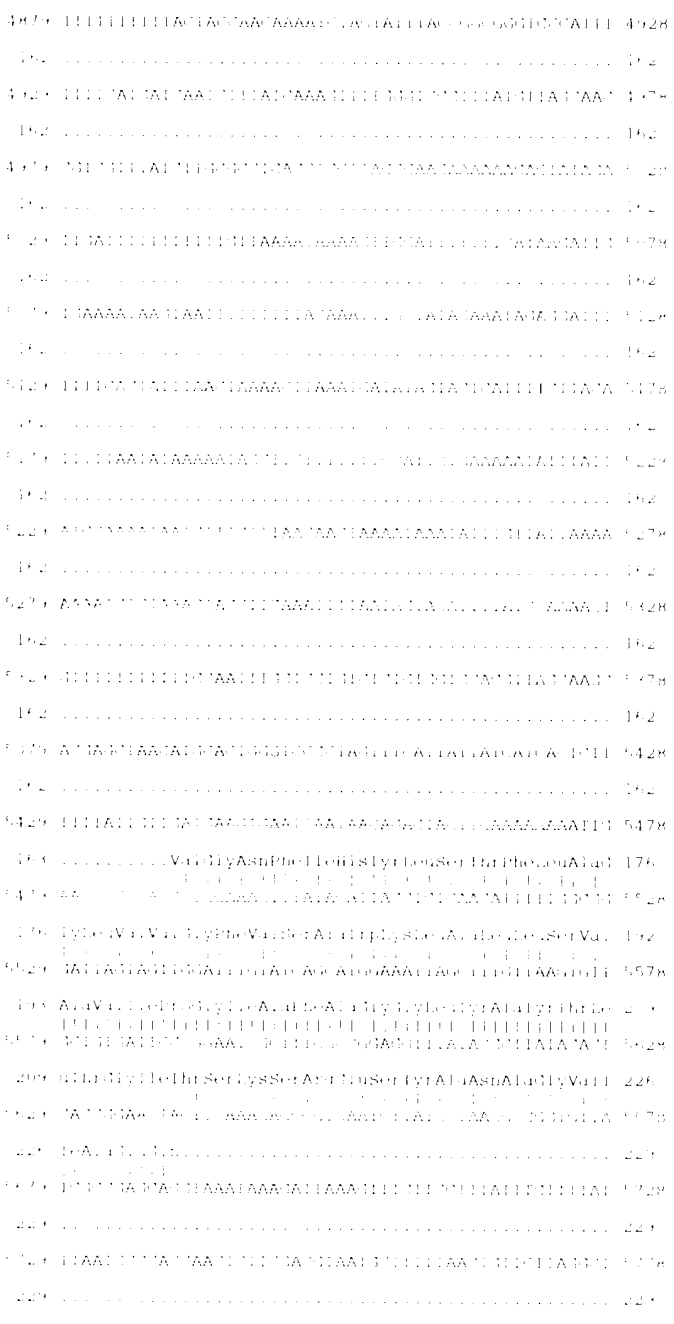


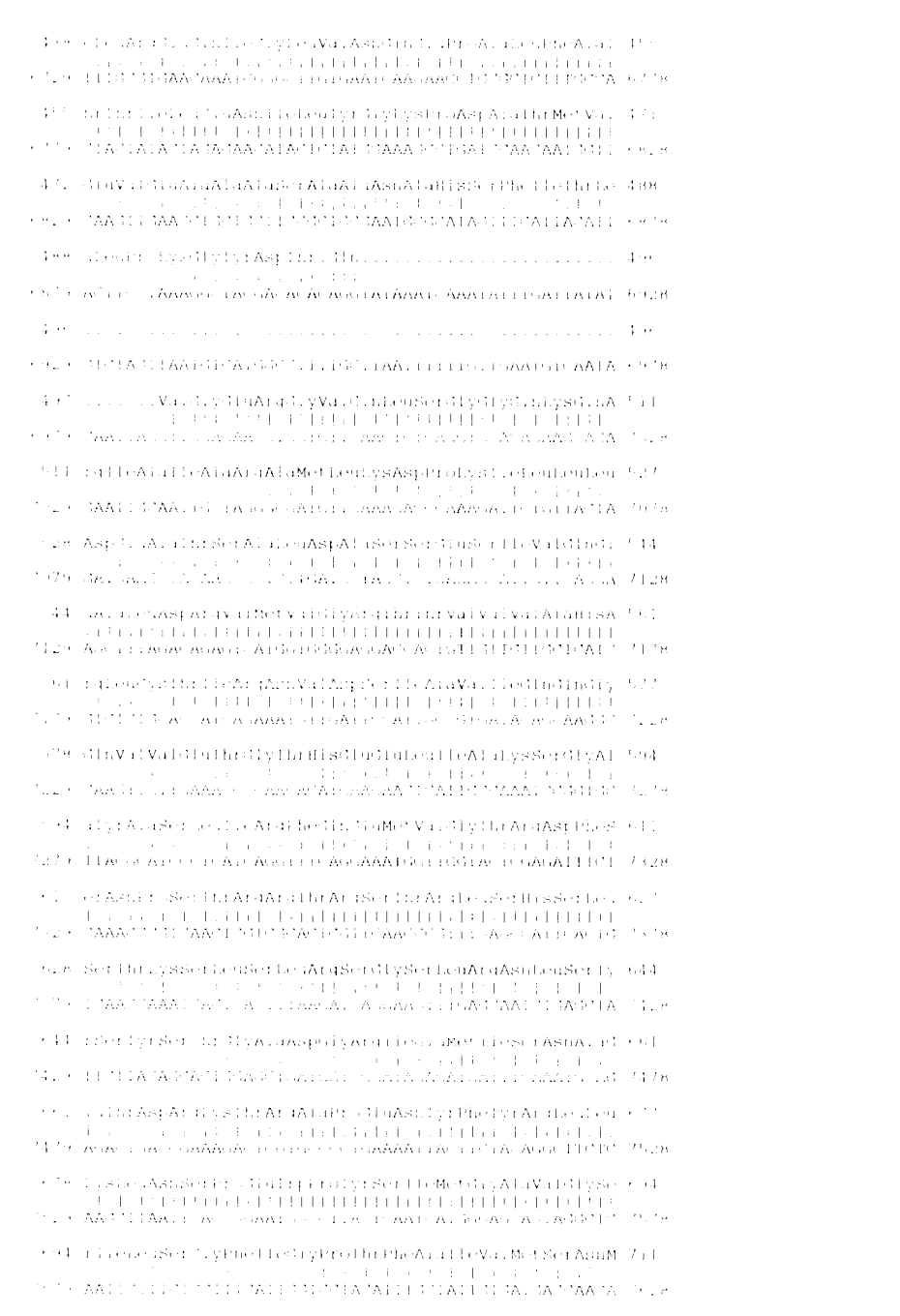


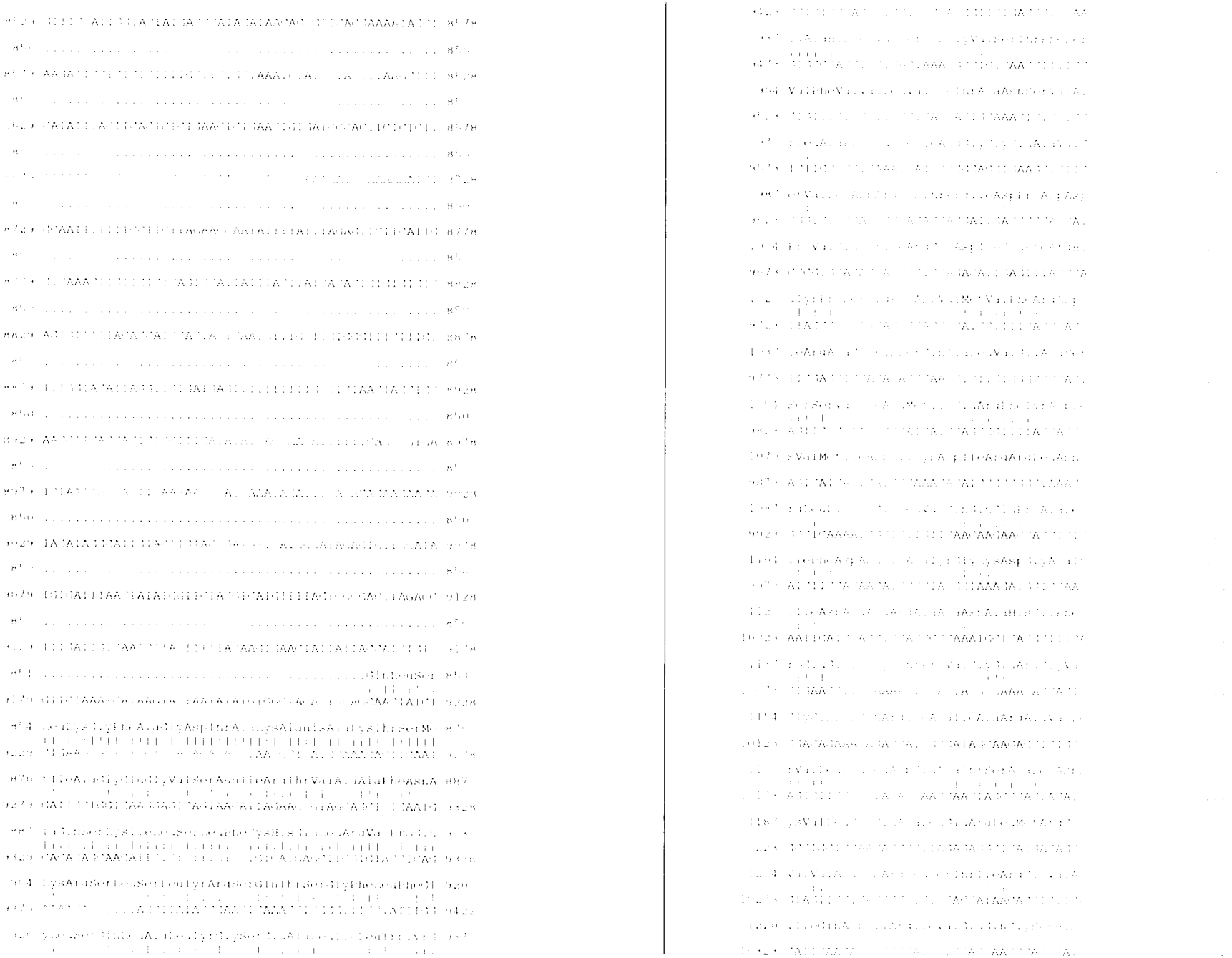


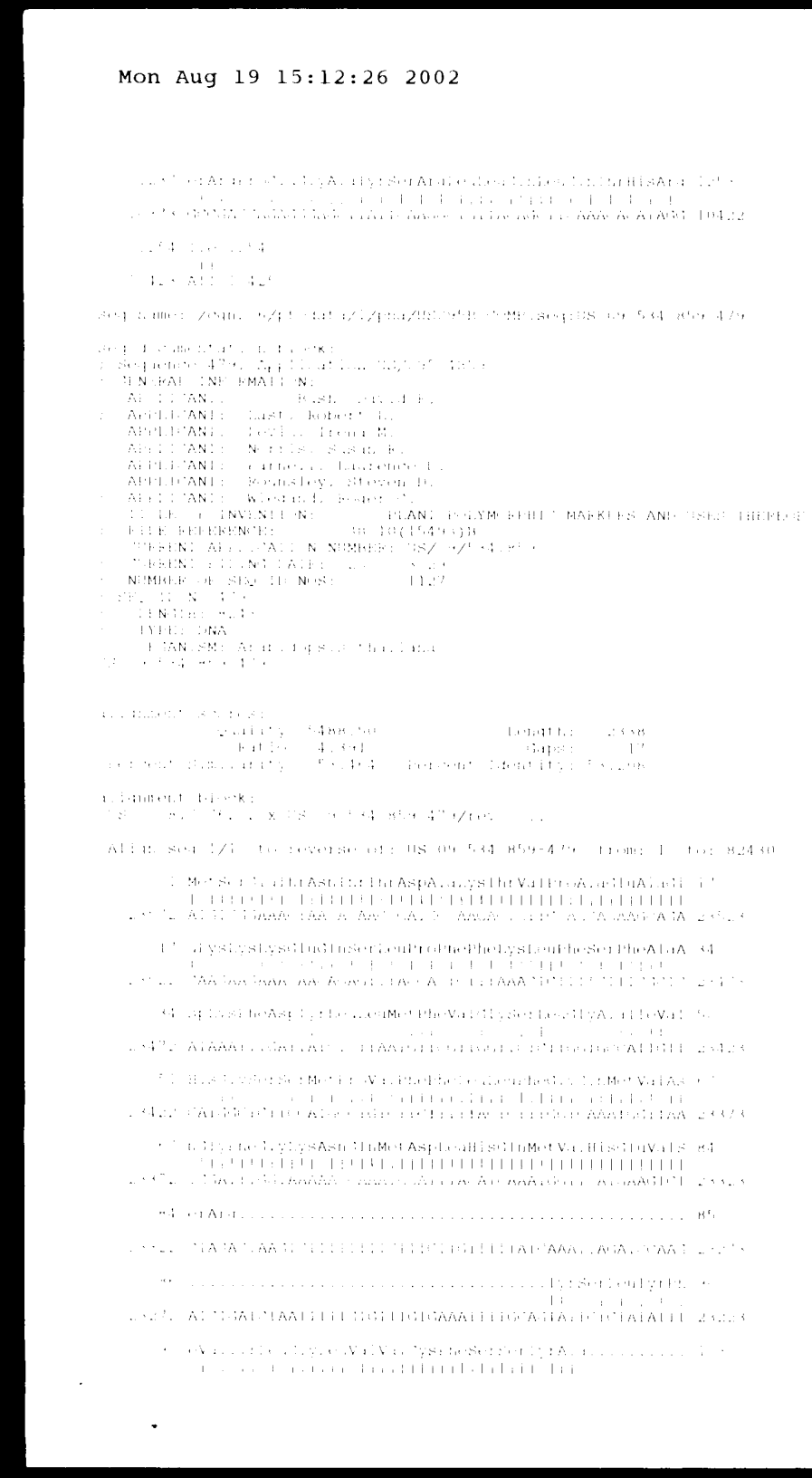


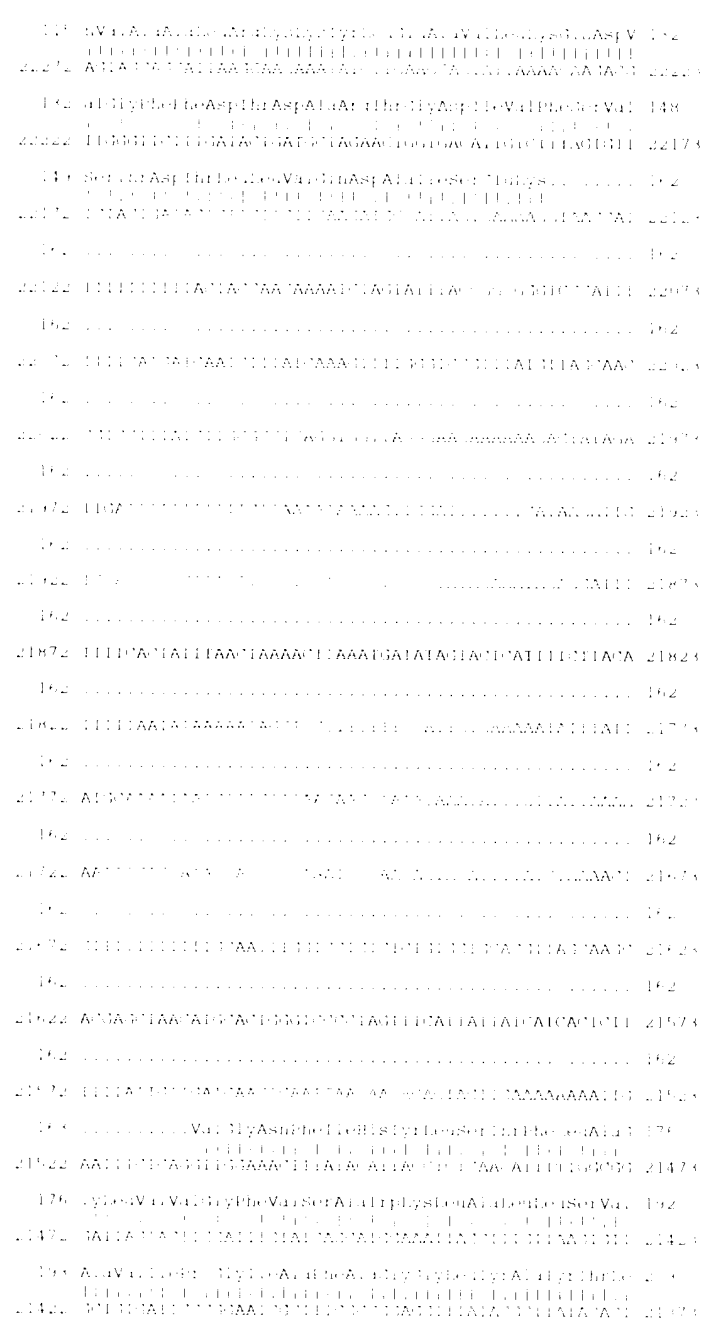


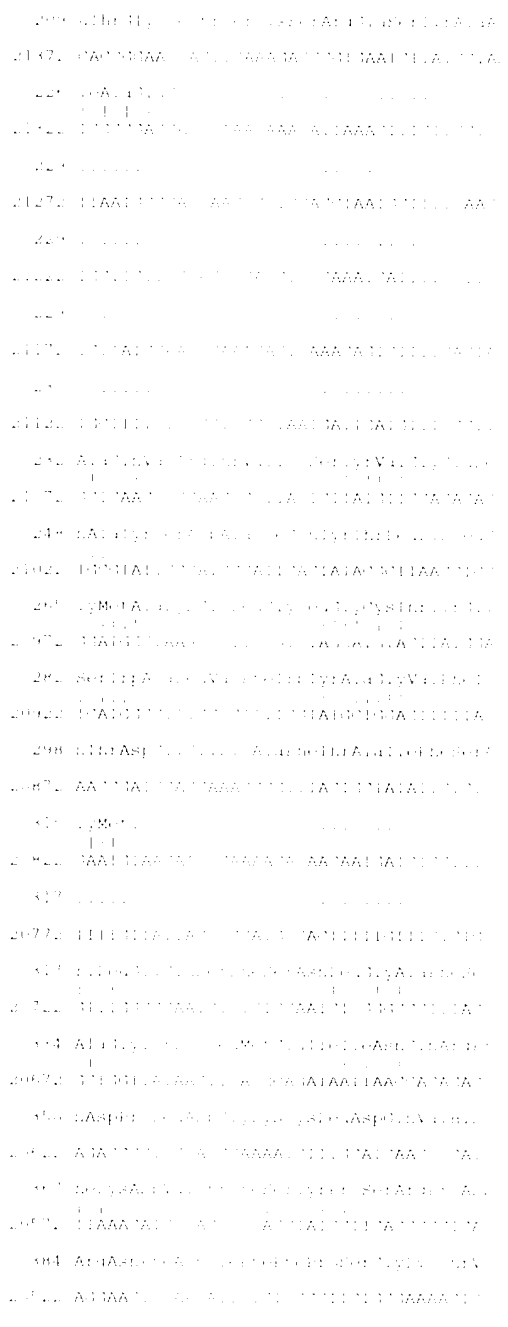


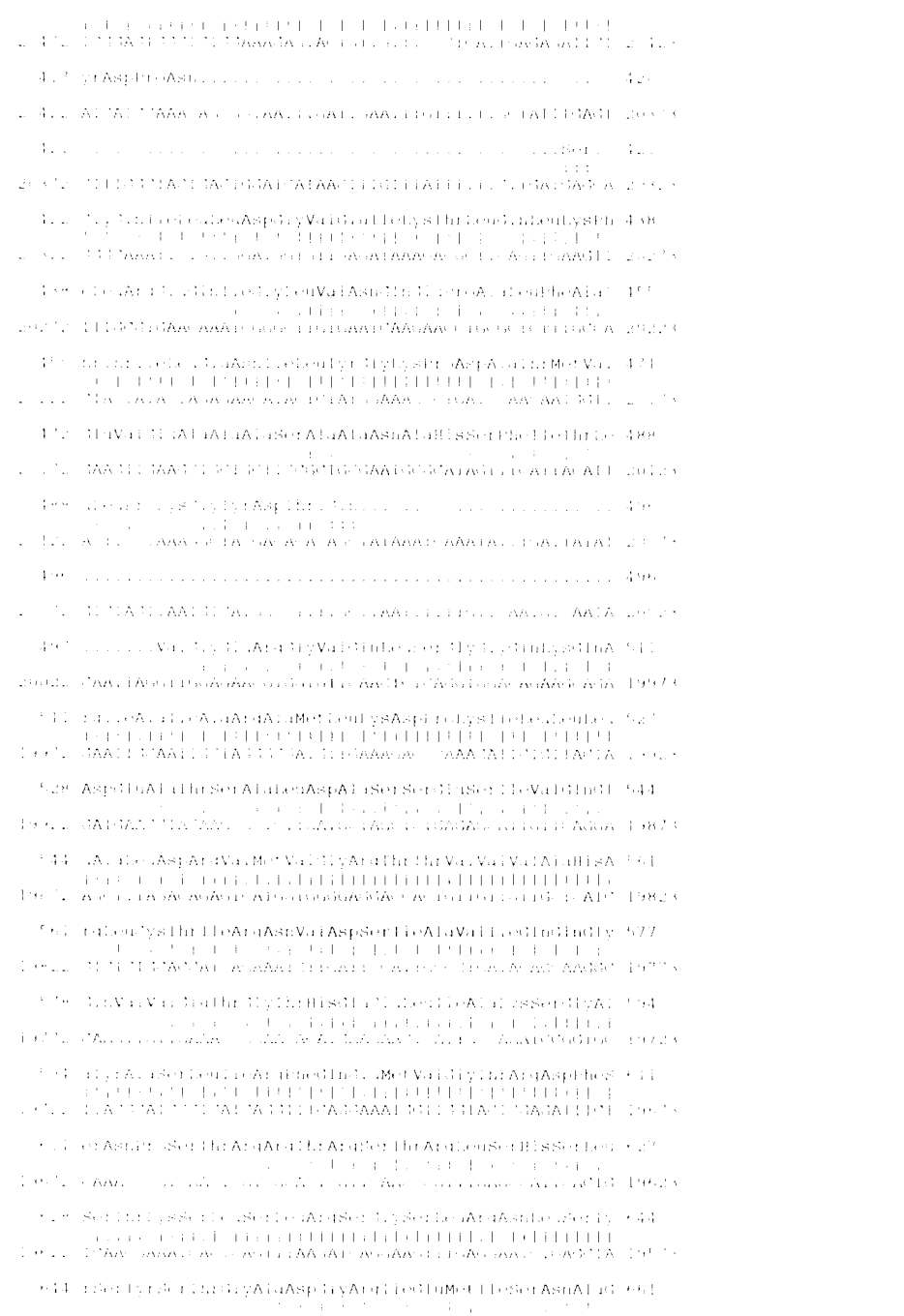


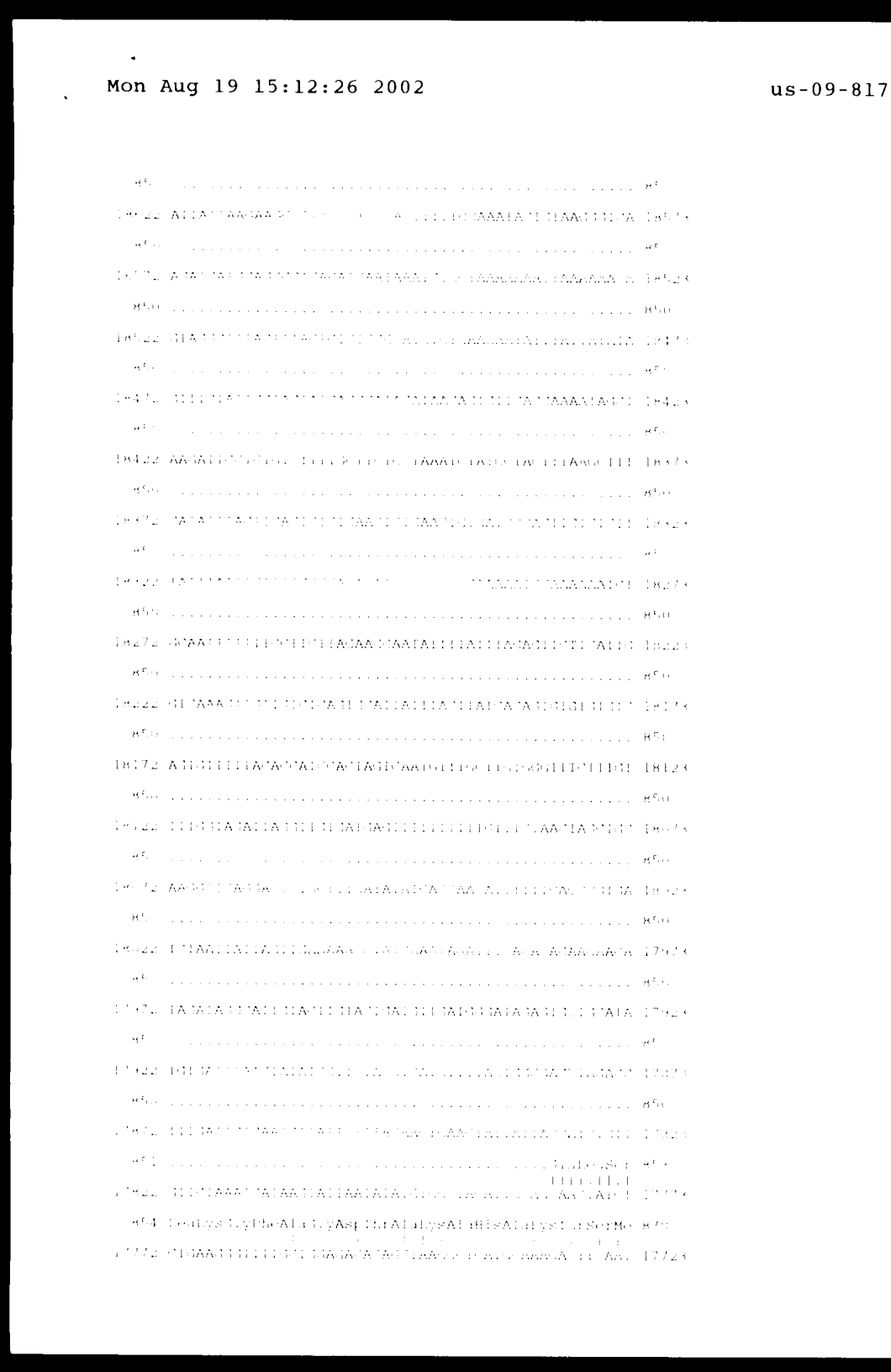


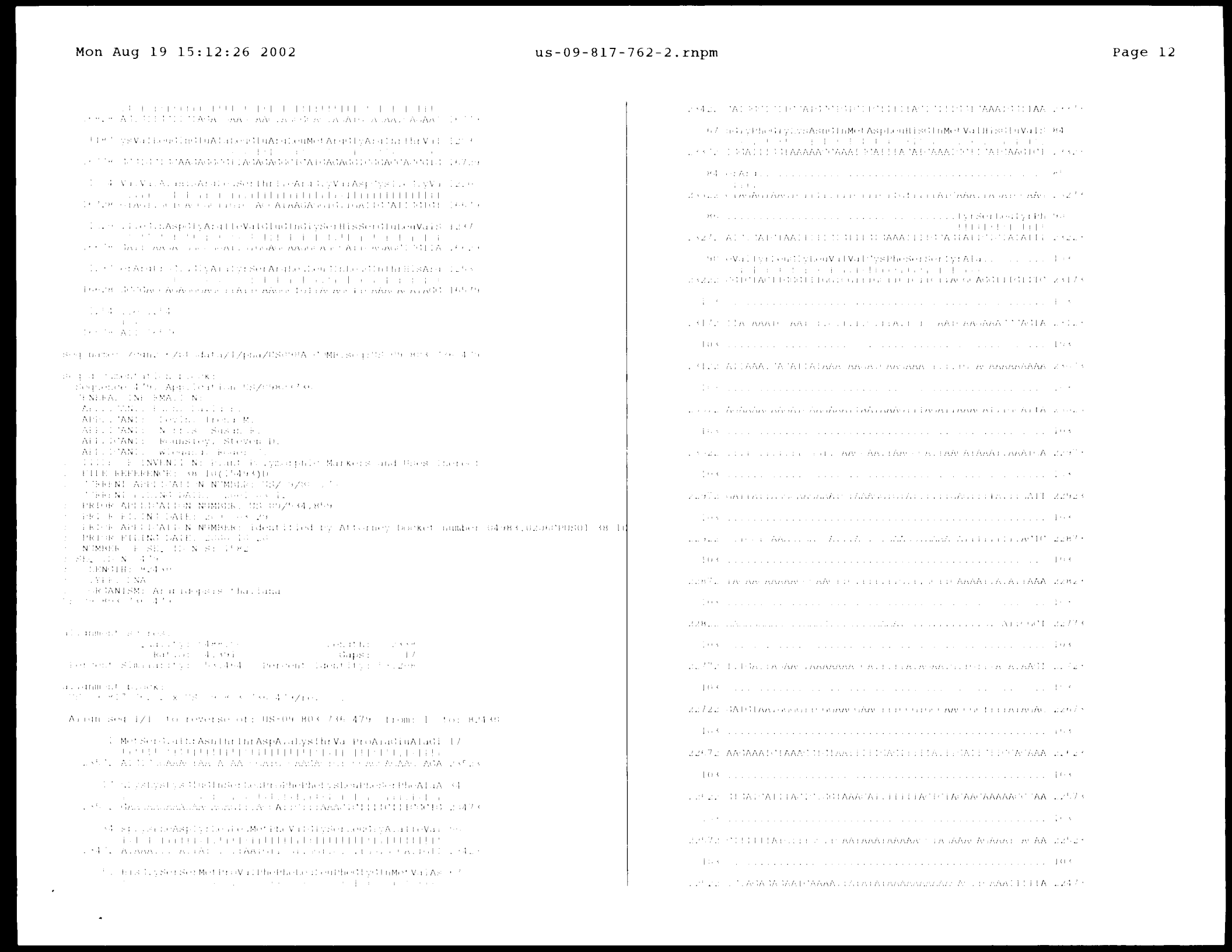


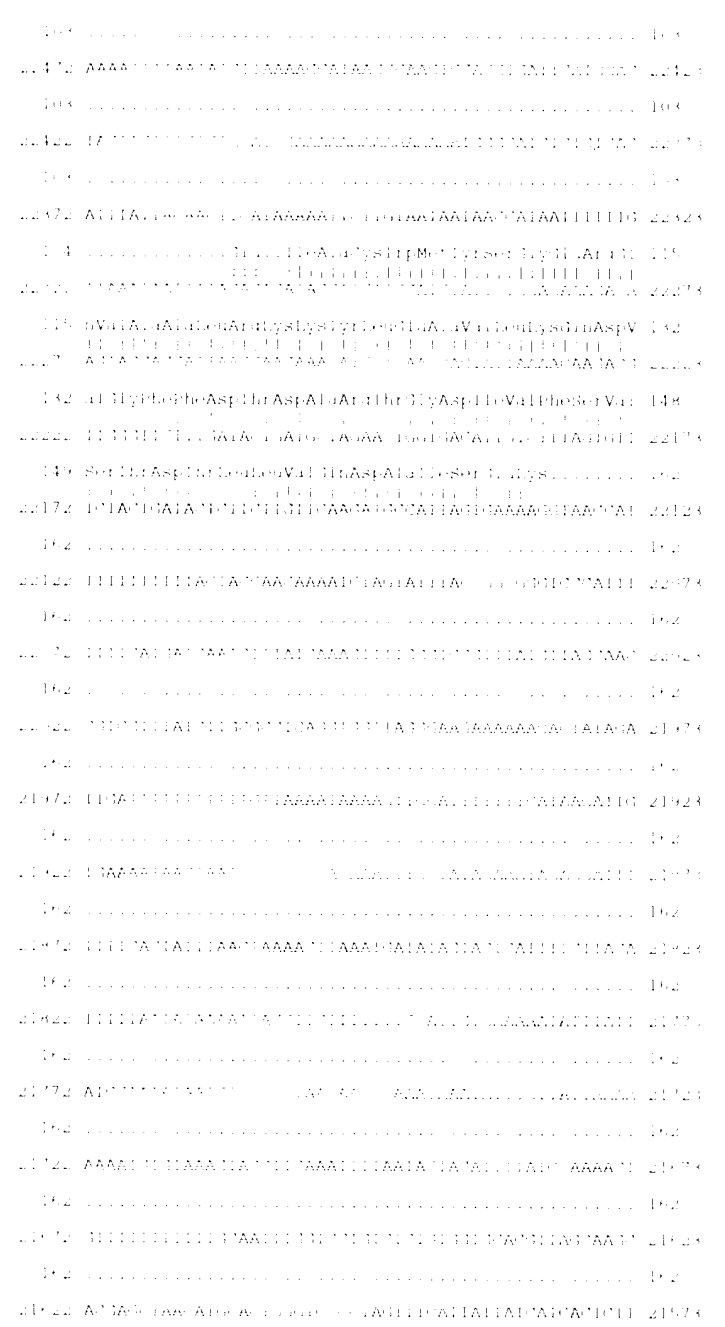


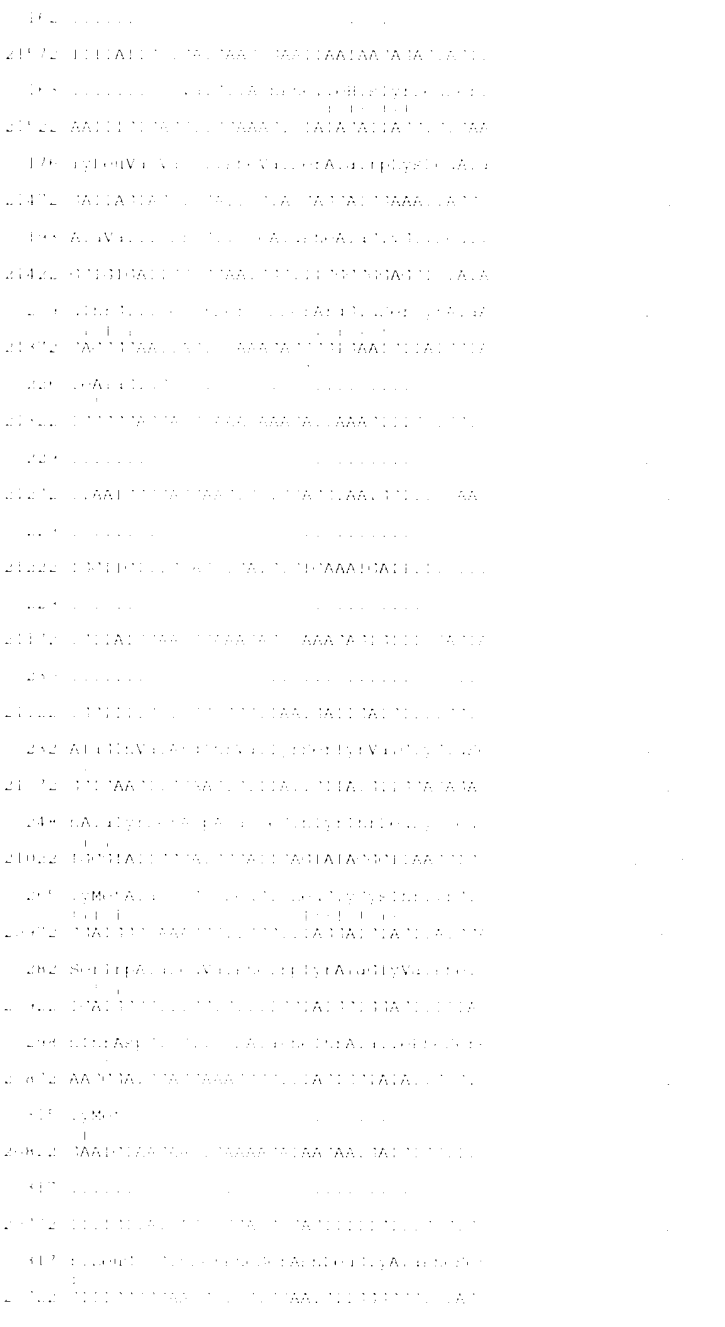


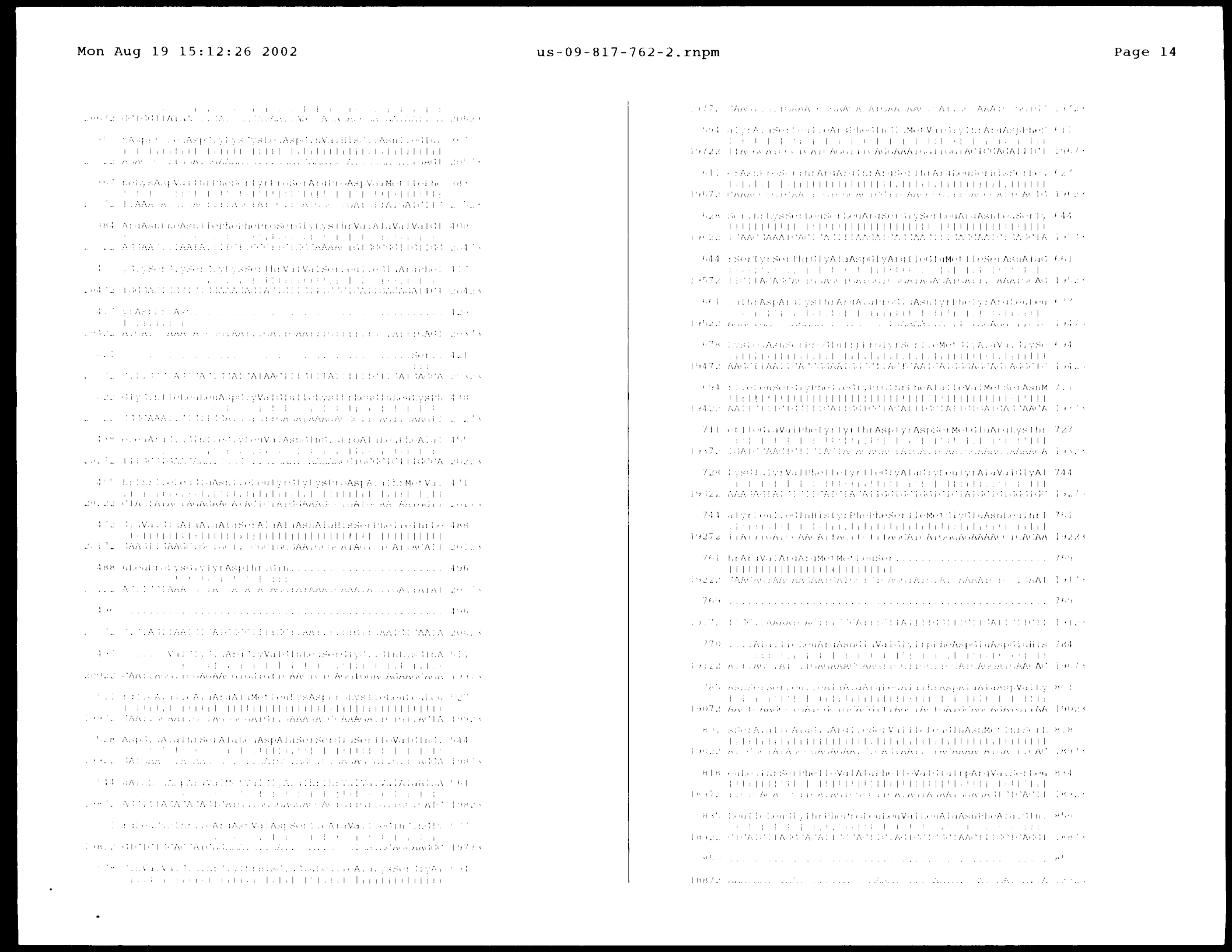


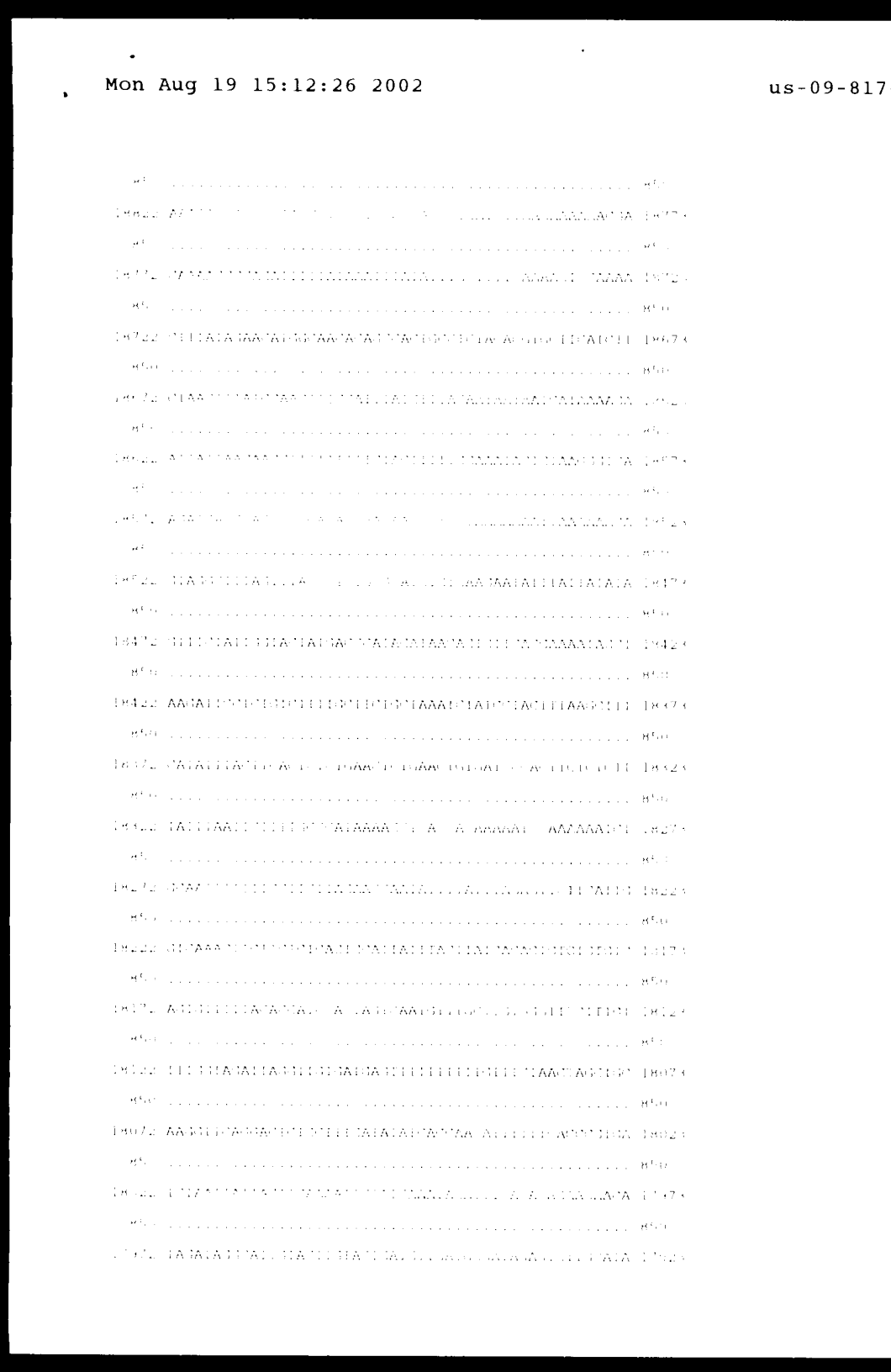


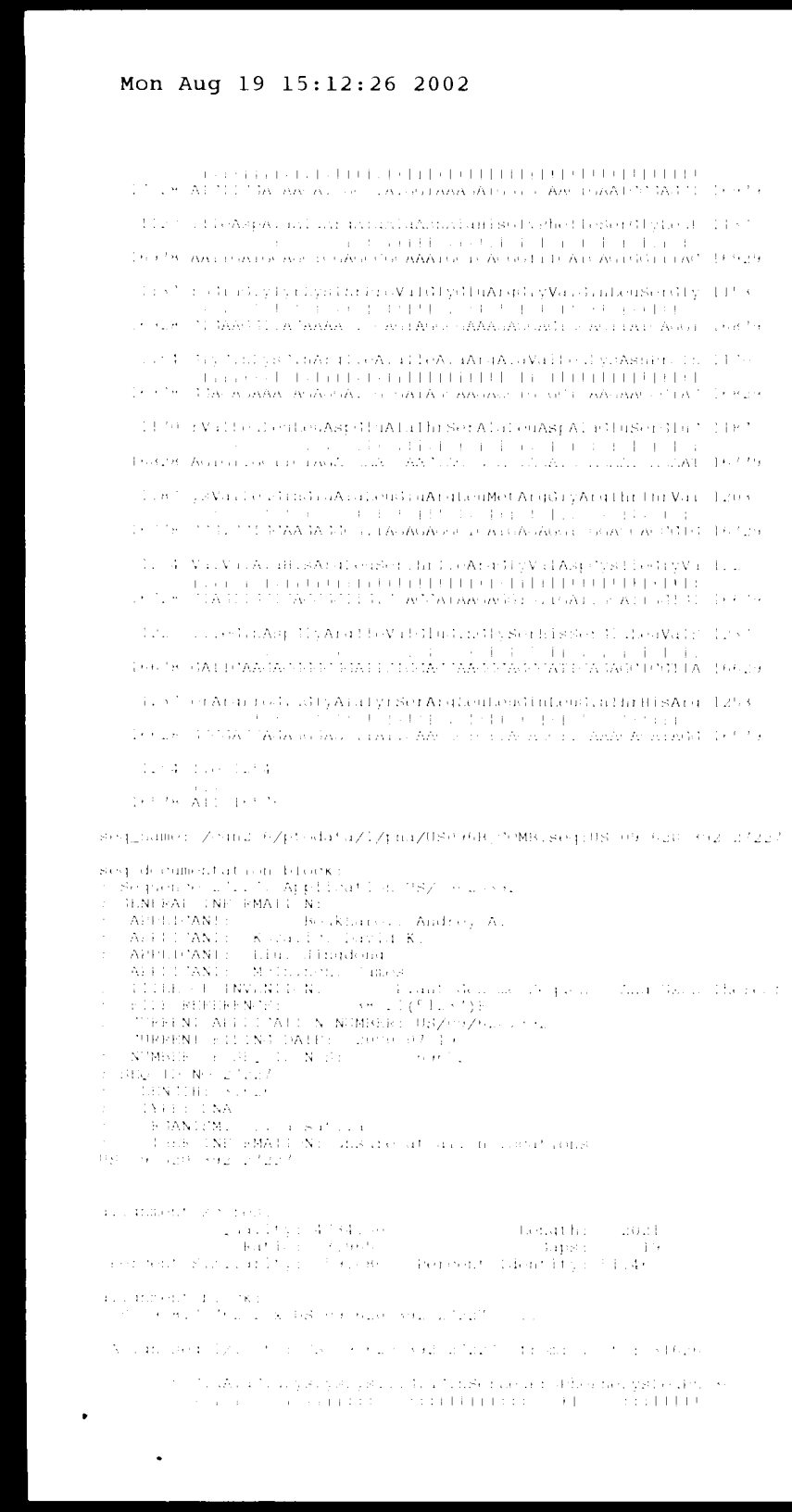












seq_number: /cgn2_6/ptodata/2/pta/US09_817_762_2_SEQ.US

sequence_information_blocks:
- Sequence 2, Application US/09817762
- GENERAL_INFORMATION:
- APPLICANT: Beckhardt, Andrew A.
- APPLICANT: Lin, Jiandong
- APPLICANT: Mitchell, Laura
- TITLE OF INVENTION: Human Secreted Proteins, and Polynucleotides Encoding Same
- FILE REFERENCE: 9[illegible]
- CURRENT APPLICATION NUMBER: US/09/817,762
- CURRENT FILING DATE: 2000-07-19
- NUMBER OF SEQ IDS: [illegible]
- SOFTWARE: PatentIn Ver. 2.1
- GENERAL INFORMATION: [illegible]

US-09-817-762-2

Query Match 100.0%; Score 2121; DB 4; Length 2121;
Best Local Similarity 100.0%; Pred. No. 0.0; Matches 2121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MetLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 19
Db 1 MetLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 19

Qy 20 ProLeuLeuLeuAlaLeuLeuAlaProGlyAlaAlaLeuAlaAlaProIleAsnLeuVal 39
Db 20 ProLeuLeuLeuAlaLeuLeuAlaProGlyAlaAlaLeuAlaAlaProIleAsnLeuVal 39

Qy 40 CysAlaAspArgGluThrProThrGlyGluAsnLeuTyrPheGlnSerGlyGlnCysLys 59
Db 40 CysAlaAspArgGluThrProThrGlyGluAsnLeuTyrPheGlnSerGlyGlnCysLys 59

Qy 60 ThrSerValAspLysIleAlaLysIleGluLysValGluMetMetSerAspMetGlnMet 79
Db 60 ThrSerValAspLysIleAlaLysIleGluLysValGluMetMetSerAspMetGlnMet 79

Qy 80 GluAspGluGluTyrSerLeuSerThrLysMetLysLysArgHisGlnGluLeuIleGln 99
Db 80 GluAspGluGluTyrSerLeuSerThrLysMetLysLysArgHisGlnGluLeuIleGln 99

Qy 100 ValLysGlnLysLeuGlnAspLeuGluAsnGluLeuAlaGlyTyrLysGluAlaAspLys 119
Db 100 ValLysGlnLysLeuGlnAspLeuGluAsnGluLeuAlaGlyTyrLysGluAlaAspLys 119

Qy 120 AlaLysLeuAlaLeuGlyAsnMetIleAsnValLeuLysGlnLeuProLysLysAsnLeu 139
Db 120 AlaLysLeuAlaLeuGlyAsnMetIleAsnValLeuLysGlnLeuProLysLysAsnLeu 139

Qy 140 AspGluIleLysProCysValSerLysAlaGlnAlaValSerHisTrpGlnGluAspPro 159
Db 140 AspGluIleLysProCysValSerLysAlaGlnAlaValSerHisTrpGlnGluAspPro 159

Qy 160 AspTyrLeuGlyLeuAlaThrGlnGlyArgGlnThrAspGlnLeuAspArgAspMetAla 179
Db 160 AspTyrLeuGlyLeuAlaThrGlnGlyArgGlnThrAspGlnLeuAspArgAspMetAla 179

Qy 180 AspGlnLeuAlaAlaLeuAsnLeuArgGlyMetAlaTyrSerIleTyrLysLeuLysSer 199
Db 180 AspGlnLeuAlaAlaLeuAsnLeuArgGlyMetAlaTyrSerIleTyrLysLeuLysSer 199

Qy 200 AsnAlaLysThrIleGluGluIleLeuAlaAlaLeuLysLysLeuLeuGluGlnMetPro 219
Db 200 AsnAlaLysThrIleGluGluIleLeuAlaAlaLeuLysLysLeuLeuGluGlnMetPro 219

Qy 220 LeuGluAspLeuGlnGlyHisLeuLeuCysLeuGluThrAlaLeuThrAsnThrThrVal 239
Db 220 LeuGluAspLeuGlnGlyHisLeuLeuCysLeuGluThrAlaLeuThrAsnThrThrVal 239

Qy 240 ValThrThrPheGlnLysGlnLeuPheTyrValArgAsnAsnLeuGluArgMetAlaHis 259
Db 240 ValThrThrPheGlnLysGlnLeuPheTyrValArgAsnAsnLeuGluArgMetAlaHis 259

Qy 260 AspIleGlyAsnSerIleGlnGlyIleAsnGlyLeuLeuSerLeuLeuGluGluMetSer 279
Db 260 AspIleGlyAsnSerIleGlnGlyIleAsnGlyLeuLeuSerLeuLeuGluGluMetSer 279

Qy 280 MetArgMetProProGluLeuArgSerGluLeuGluHisHisLeuLysGlnAsnIleVal 299
Db 280 MetArgMetProProGluLeuArgSerGluLeuGluHisHisLeuLysGlnAsnIleVal 299

Qy 300 AsnLeuSerGlnSerProAlaGlnHisIleGlnGluValAlaAlaGlnValGlnSerSer 319
Db 300 AsnLeuSerGlnSerProAlaGlnHisIleGlnGluValAlaAlaGlnValGlnSerSer 319

Qy 320 TrpGluTrpGlnArgLeuLeuGlnAsnLeuGlnGlyArgLeuGluArgTyrSerAla 339
Db 320 TrpGluTrpGlnArgLeuLeuGlnAsnLeuGlnGlyArgLeuGluArgTyrSerAla 339

Qy 340 LeuLeuGluProGlyAlaArgSerArgLeuLeuSerIleLeuGlnAlaArgProPheLeu 359
Db 340 LeuLeuGluProGlyAlaArgSerArgLeuLeuSerIleLeuGlnAlaArgProPheLeu 359

Qy 360 GluTrpMetGlyValLeuGluHisGlyLysGlnSerLeuMetLeuGlnAspValMet 379
Db 360 GluTrpMetGlyValLeuGluHisGlyLysGlnSerLeuMetLeuGlnAspValMet 379

Qy 380 GluGlyAsnTrpIleLeuIleLysAlaProLysGlnGluGlyGlyGluGlyAsnGly 399
Db 380 GluGlyAsnTrpIleLeuIleLysAlaProLysGlnGluGlyGlyGluGlyAsnGly 399

Qy 400 LeuLeuLysGluLeuGlnAlaGluTyrIleLeuArgProGlyProValHisGlyLeu 419
Db 400 LeuLeuLysGluLeuGlnAlaGluTyrIleLeuArgProGlyProValHisGlyLeu 419

Qy 420 AlaAsnLeuArgLeuValValGluAsnLeuGlnThrSerGlyLeuTyrGlyLeuSer 439
Db 420 AlaAsnLeuArgLeuValValGluAsnLeuGlnThrSerGlyLeuTyrGlyLeuSer 439

Qy 440 LeuArgLysAlaGlyGluProLeuGlnLeuArgLeuProGlyLeuProProArgLeu 459
Db 440 LeuArgLysAlaGlyGluProLeuGlnLeuArgLeuProGlyLeuProProArgLeu 459

Qy 460 AspLeuSerHisLeuThrGlyPheLysLeuProProLeuMetSerLeuArgLeuPro 479
Db 460 AspLeuSerHisLeuThrGlyPheLysLeuProProLeuMetSerLeuArgLeuPro 479

Qy 480 AspLeuArgLeuProHisLeuGlnValLeuAlaLeuAspLeuAlaThrProGlyIle 499
Db 480 AspLeuArgLeuProHisLeuGlnValLeuAlaLeuAspLeuAlaThrProGlyIle 499

Qy 500 SerAspLeuSerPheLeuGluAlaAlaAlaSerValLeuAlaLeuLeuSerGlnVal 519
Db 500 SerAspLeuSerPheLeuGluAlaAlaAlaSerValLeuAlaLeuLeuSerGlnVal 519

Qy 520 IleIleLeuGlyLeuSerLeuValValGlyGlnValValProLeuSerGluGluIle 539
Db 520 IleIleLeuGlyLeuSerLeuValValGlyGlnValValProLeuSerGluGluIle 539

Qy 540 LysCysValSerSerGluValProValIleGlnGlnAsnProLysThrMetGlnGlu 559
Db 540 LysCysValSerSerGluValProValIleGlnGlnAsnProLysThrMetGlnGlu 559

Qy 560 ThrLeuPheGlnLysLeuIleLysTrpLeuAspLysAsnGlnGlnValLeuProVal 579
Db 560 ThrLeuPheGlnLysLeuIleLysTrpLeuAspLysAsnGlnGlnValLeuProVal 579

Qy 580 ThrIleThrAlaThrValGlyMetAlaLeuThrSerLeuHisSerGluIleAsnPro 599
Db 580 ThrIleThrAlaThrValGlyMetAlaLeuThrSerLeuHisSerGluIleAsnPro 599

Qy 600 LeuGlnAsnCysValPheProGlnGluGlnValGlnGlyLeuArgGluGlnGluPhe 619
Db 600 LeuGlnAsnCysValPheProGlnGluGlnValGlnGlyLeuArgGluGlnGluPhe 619

Qy 620 LeuLysLysAsnLysGluGluLeuLysGluCysPheGlnGluLeuThrGlnLysLeu 639
Db 620 LeuLysLysAsnLysGluGluLeuLysGluCysPheGlnGluLeuThrGlnLysLeu 639

Qy 640 HisGlnAlaValSerGlnAlaGlnGlnGlnMetGluGlnLeuGlnGluSerLeuArg 659
Db 640 HisGlnAlaValSerGlnAlaGlnGlnGlnMetGluGlnLeuGlnGluSerLeuArg 659

Qy 660 GlnGluValAlaArgArgAlaGlnLeuLeuGlnThrGlyLeuAspValPheLeuTrp 679
Db 660 GlnGluValAlaArgArgAlaGlnLeuLeuGlnThrGlyLeuAspValPheLeuTrp 679

Qy 680 LeuAlaGluGlyMetLeuLeuGlnGlnAspLeuGlyGlnGlnLeuGlnGluCysThr 699
Db 680 LeuAlaGluGlyMetLeuLeuGlnGlnAspLeuGlyGlnGlnLeuGlnGluCysThr 699

Qy 700 GlnLeuLysGlnArgLeuGlnGlnGluLeuGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 719
Db 700 GlnLeuLysGlnArgLeuGlnGlnGluLeuGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 719

Qy 720 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 739
Db 720 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 739

Qy 740 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 759
Db 740 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 759

Qy 760 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 779
Db 760 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 779

Qy 780 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 799
Db 780 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 799

Qy 800 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 819
Db 800 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 819

Qy 820 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 839
Db 820 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 839

Qy 840 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 859
Db 840 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 859

Qy 860 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 879
Db 860 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 879

Qy 880 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 899
Db 880 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 899

Qy 900 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 919
Db 900 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 919

Qy 920 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 939
Db 920 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 939

Qy 940 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 959
Db 940 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 959

Qy 960 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 979
Db 960 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 979

Qy 980 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 999
Db 980 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 999

Qy 1000 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1019
Db 1000 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1019

Qy 1020 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1039
Db 1020 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1039

Qy 1040 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1059
Db 1040 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1059

Qy 1060 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1079
Db 1060 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1079

Qy 1080 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1099
Db 1080 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1099

Qy 1100 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1119
Db 1100 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1119

Qy 1120 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1139
Db 1120 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1139

Qy 1140 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1159
Db 1140 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1159

Qy 1160 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1179
Db 1160 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1179

Qy 1180 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1199
Db 1180 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1199

Qy 1200 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1219
Db 1200 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1219

Qy 1220 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1239
Db 1220 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1239

Qy 1240 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1259
Db 1240 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1259

Qy 1260 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1279
Db 1260 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1279

Qy 1280 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1299
Db 1280 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1299

Qy 1300 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1319
Db 1300 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1319

Qy 1320 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1339
Db 1320 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1339

Qy 1340 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1359
Db 1340 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1359

Qy 1360 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1379
Db 1360 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1379

Qy 1380 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1399
Db 1380 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1399

Qy 1400 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1419
Db 1400 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1419

Qy 1420 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1439
Db 1420 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1439

Qy 1440 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1459
Db 1440 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1459

Qy 1460 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1479
Db 1460 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1479

Qy 1480 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1499
Db 1480 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1499

Qy 1500 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1519
Db 1500 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1519

Qy 1520 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1539
Db 1520 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1539

Qy 1540 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1559
Db 1540 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1559

Qy 1560 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1579
Db 1560 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1579

Qy 1580 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1599
Db 1580 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1599

Qy 1600 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1619
Db 1600 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1619

Qy 1620 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1639
Db 1620 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1639

Qy 1640 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1659
Db 1640 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1659

Qy 1660 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1679
Db 1660 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1679

Qy 1680 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1699
Db 1680 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1699

Qy 1700 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1719
Db 1700 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1719

Qy 1720 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1739
Db 1720 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1739

Qy 1740 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1759
Db 1740 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1759

Qy 1760 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1779
Db 1760 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1779

Qy 1780 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1799
Db 1780 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1799

Qy 1800 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1819
Db 1800 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1819

Qy 1820 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1839
Db 1820 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1839

Qy 1840 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1859
Db 1840 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1859

Qy 1860 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1879
Db 1860 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1879

Qy 1880 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1899
Db 1880 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1899

Qy 1900 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1919
Db 1900 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1919

Qy 1920 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1939
Db 1920 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1939

Qy 1940 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1959
Db 1940 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1959

Qy 1960 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1979
Db 1960 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1979

Qy 1980 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1999
Db 1980 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 1999

Qy 2000 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2019
Db 2000 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2019

Qy 2020 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2039
Db 2020 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2039

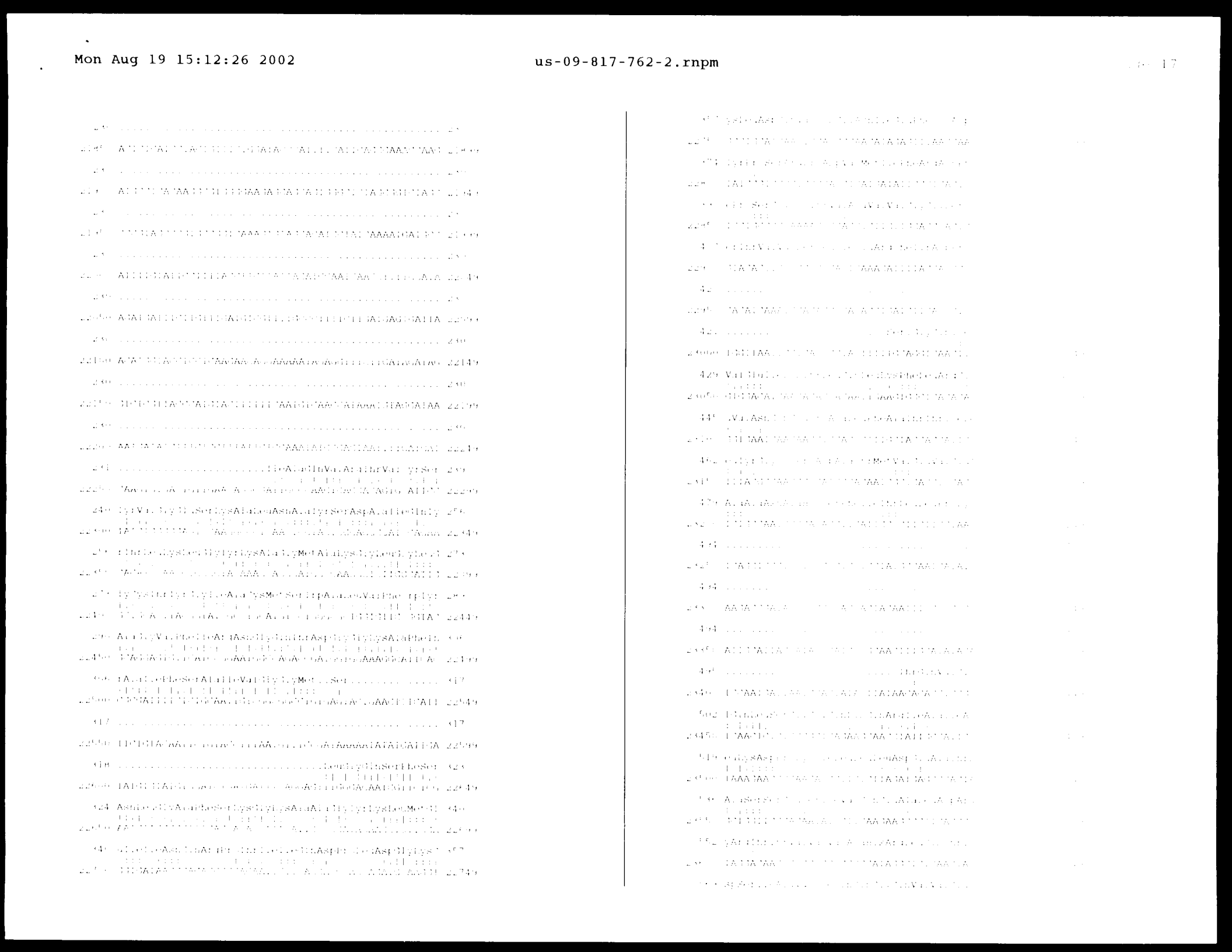
Qy 2040 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2059
Db 2040 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2059

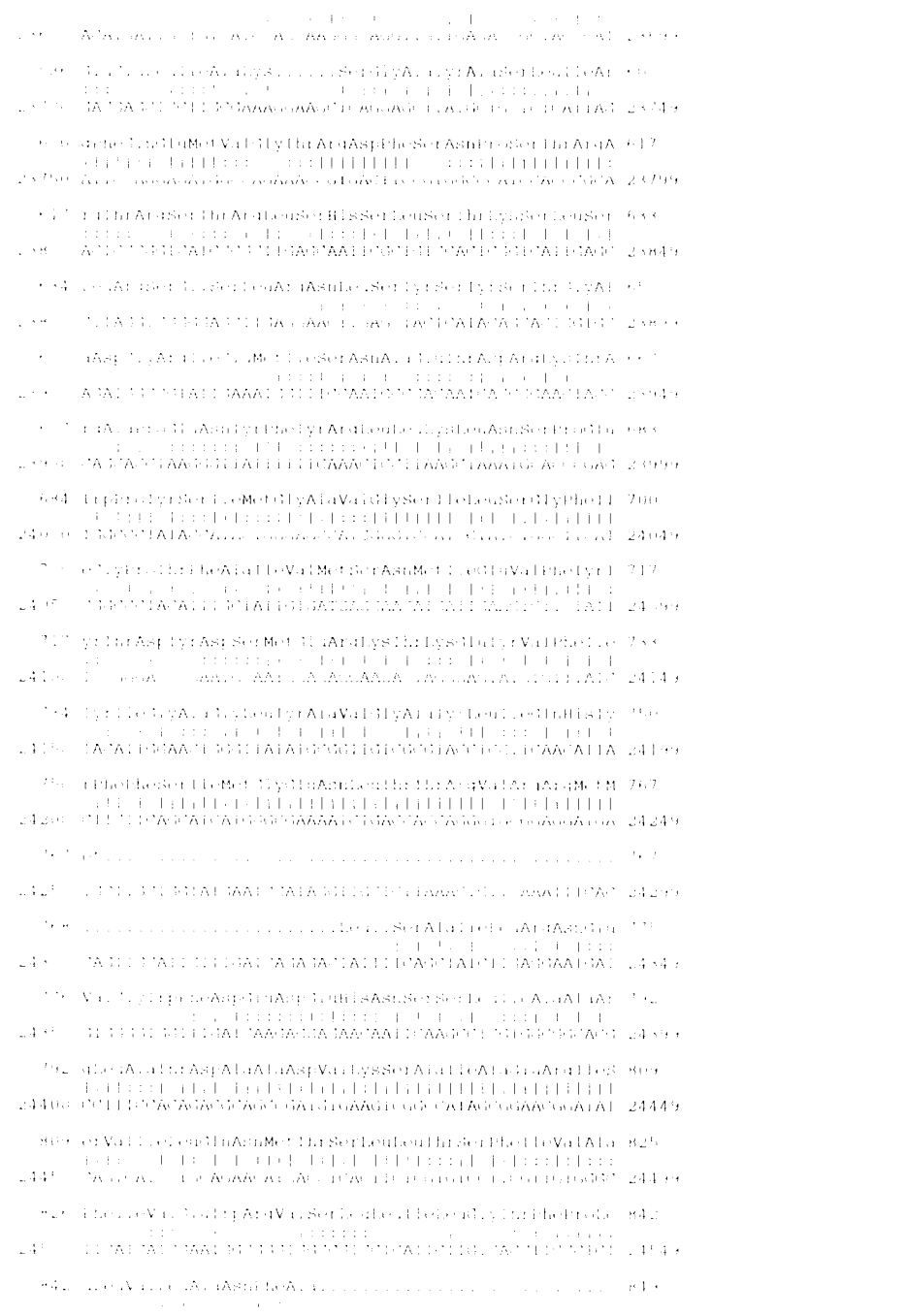
Qy 2060 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2079
Db 2060 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2079

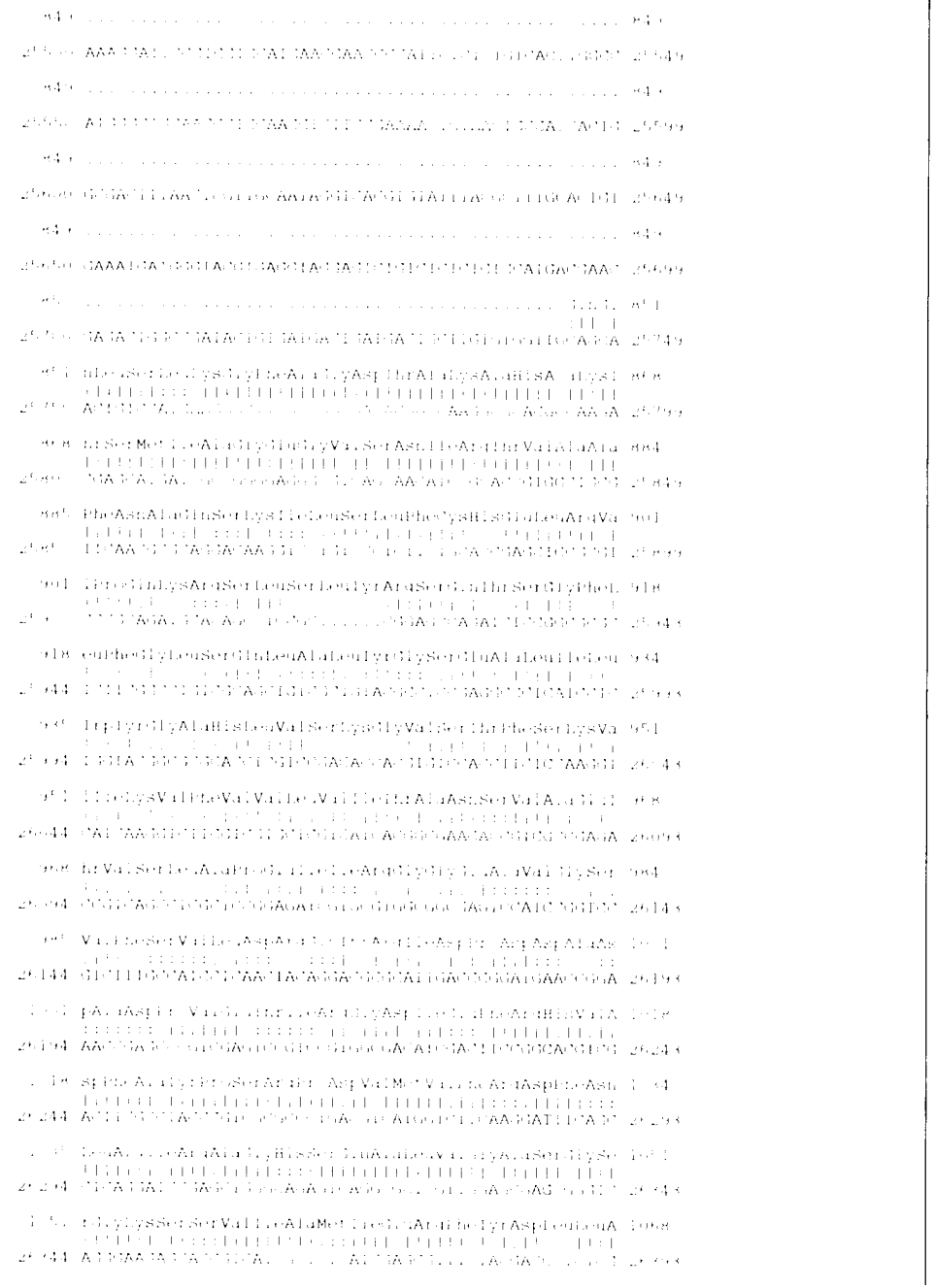
Qy 2080 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2099
Db 2080 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2099

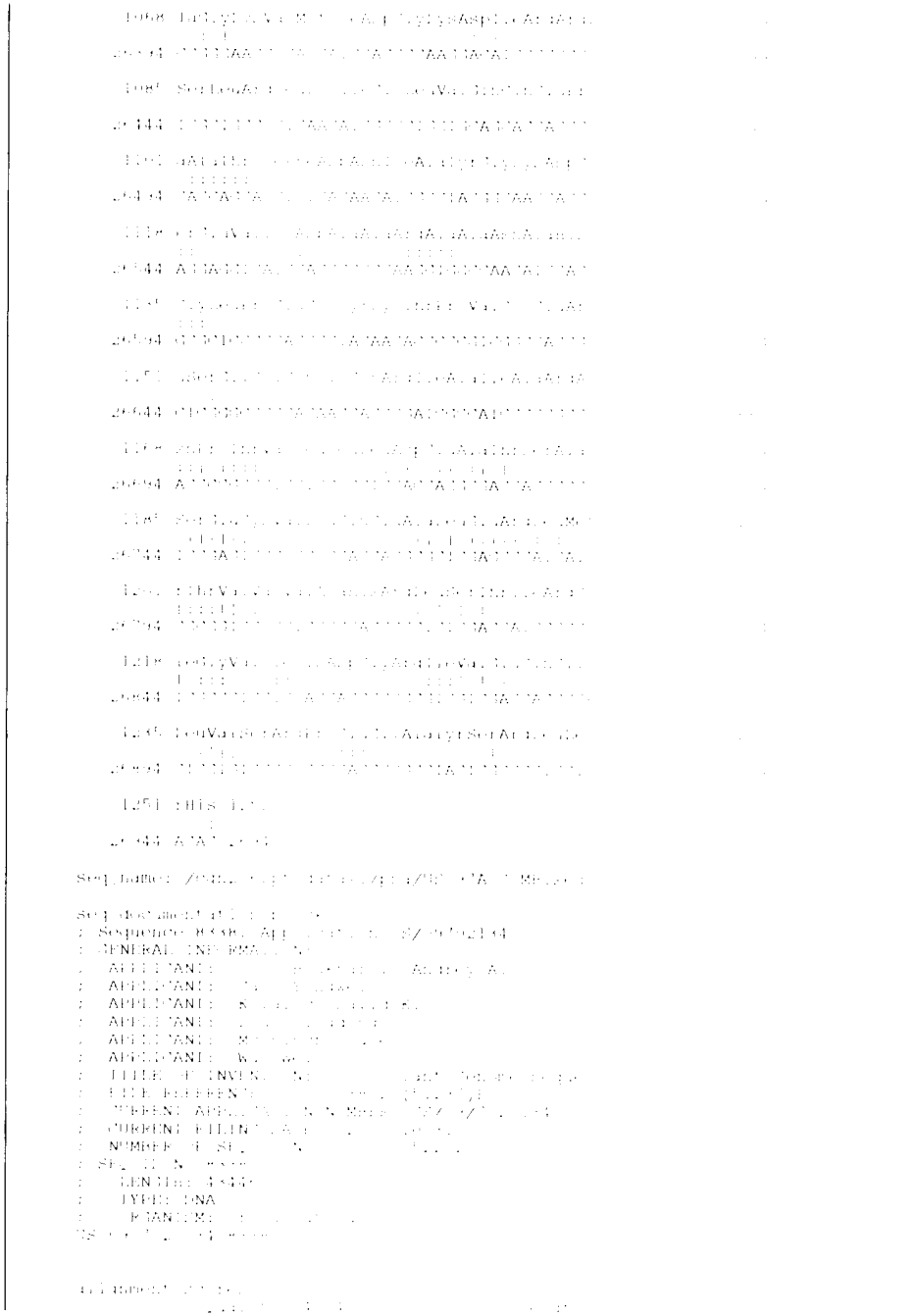
Qy 2100 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2119
Db 2100 LeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuGlnGlnGln 2119

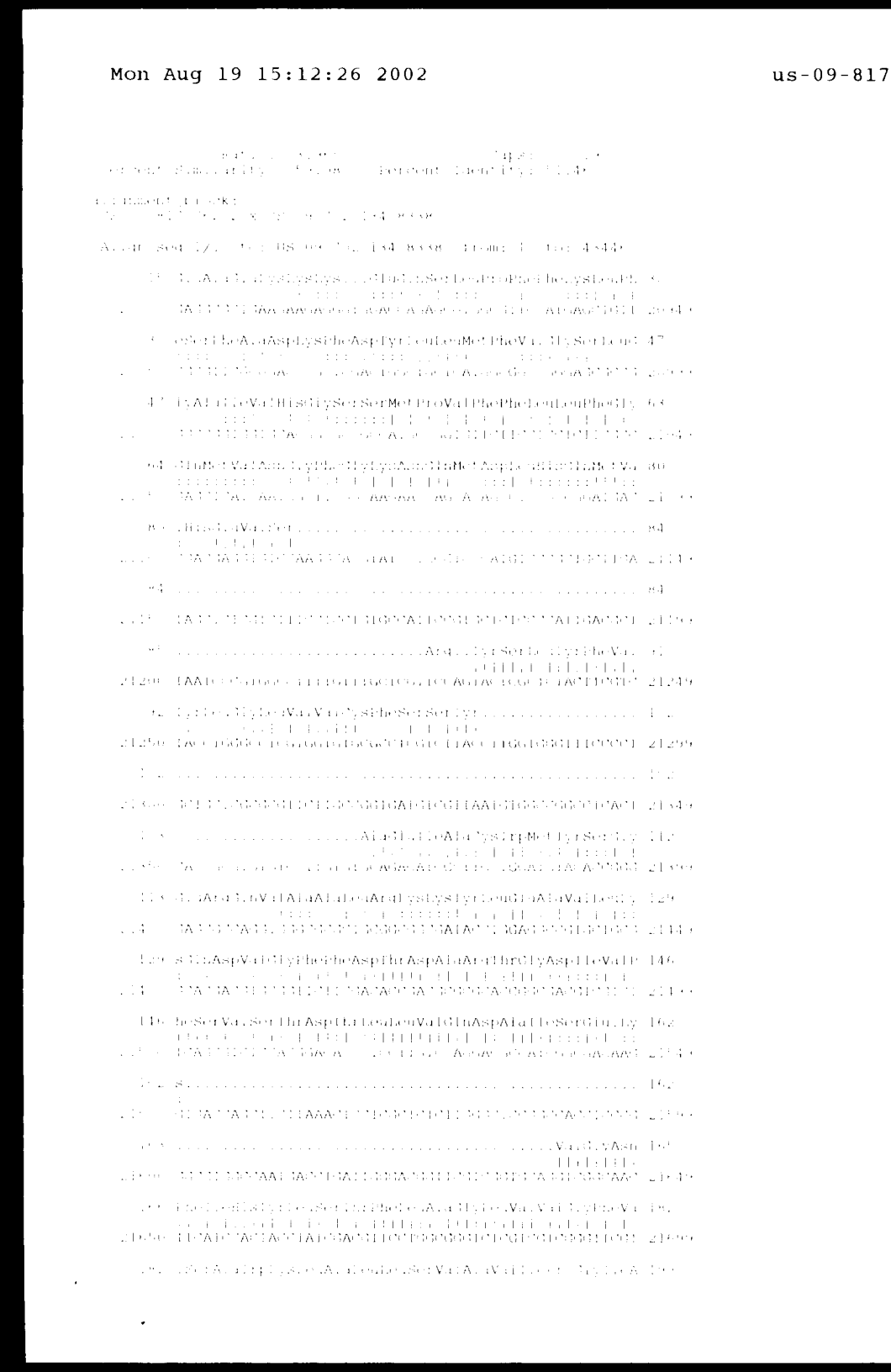
Qy 2120 LeuGln 2121
Db 2120 LeuGln 2121

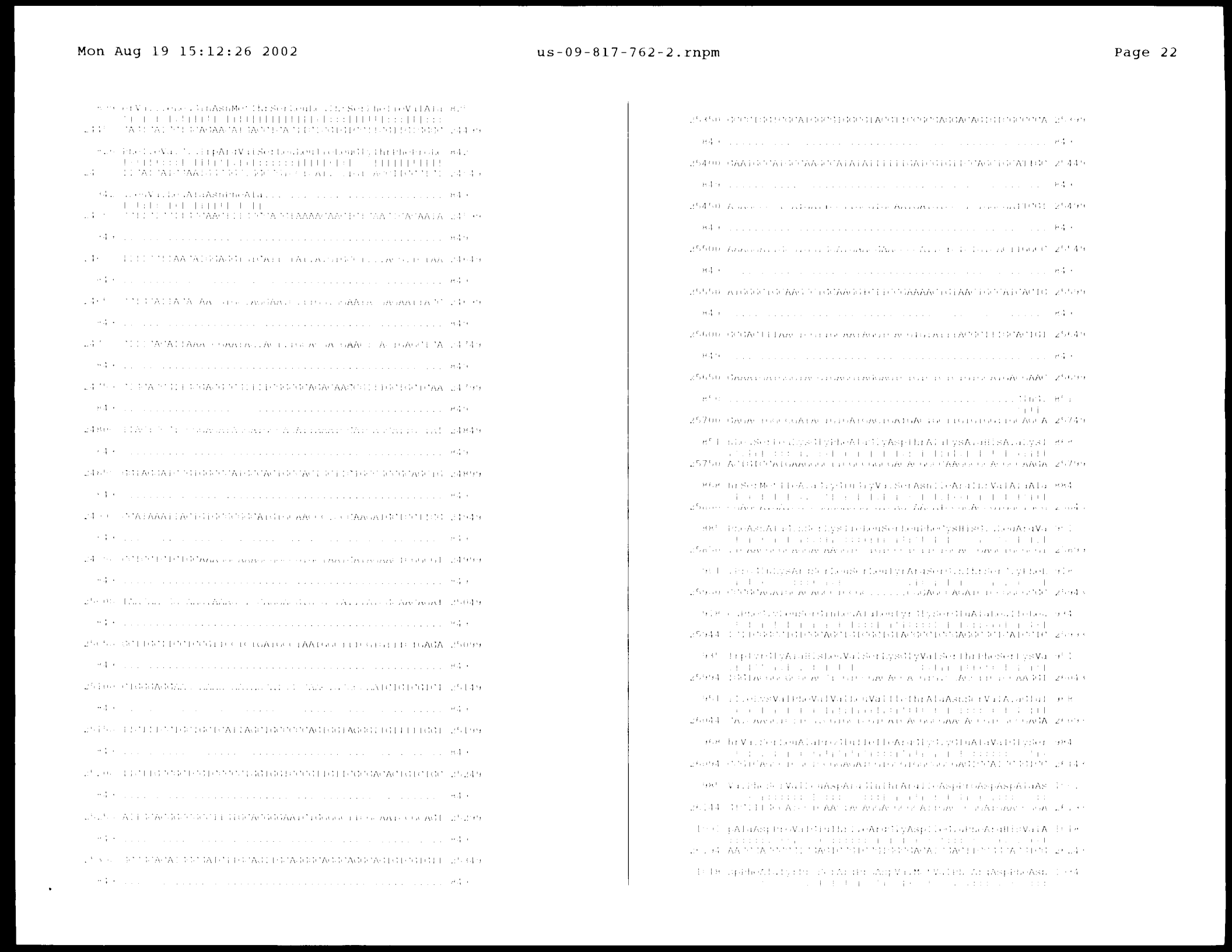


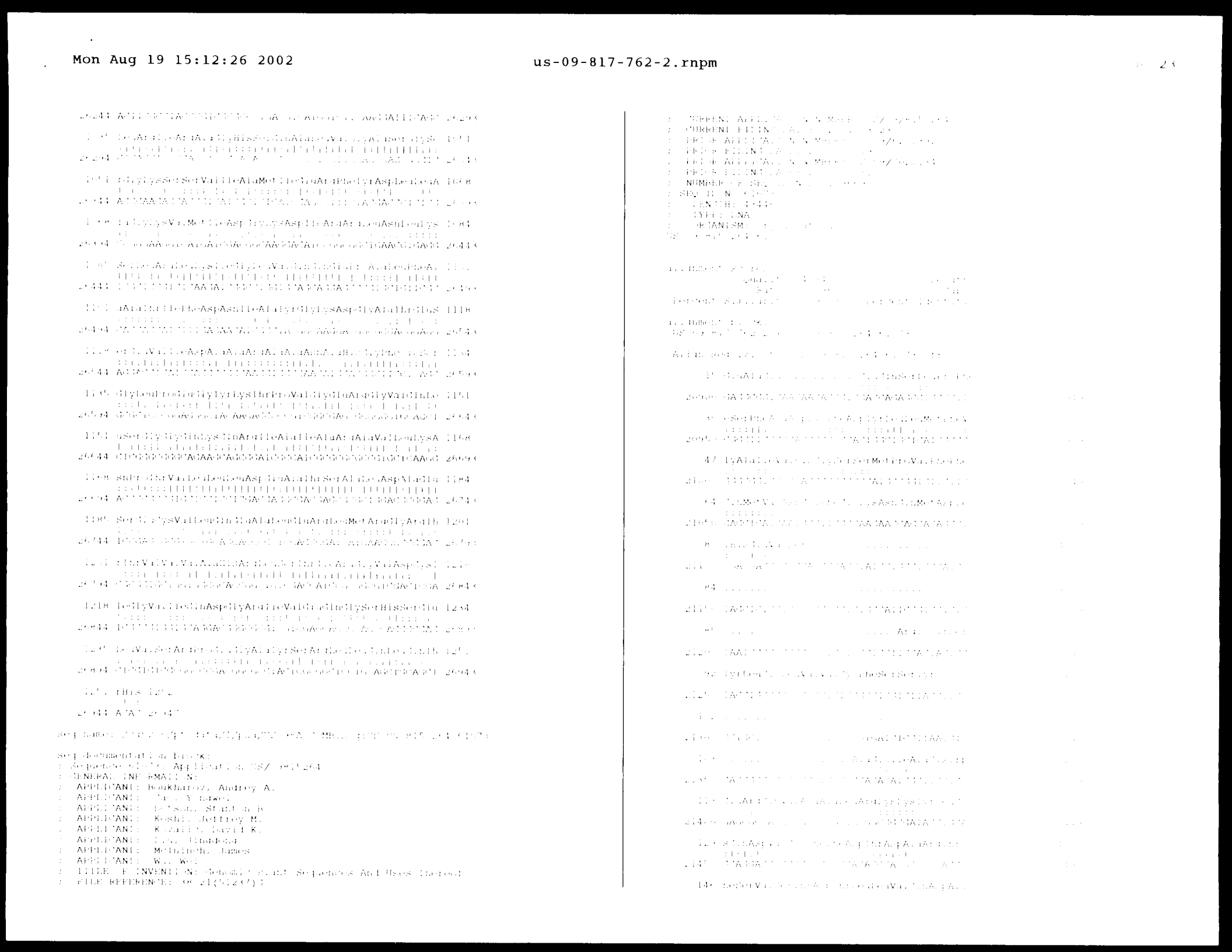


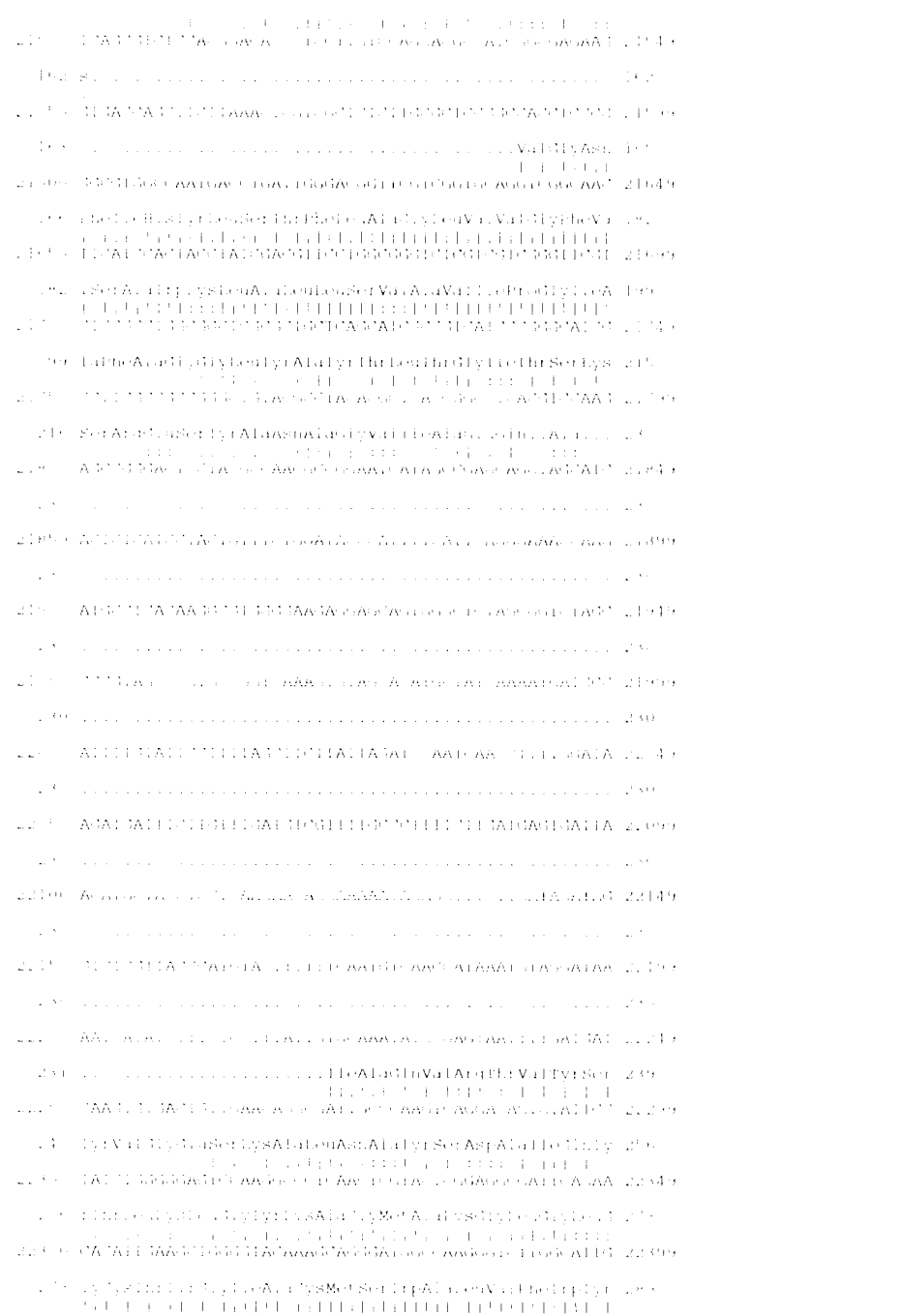


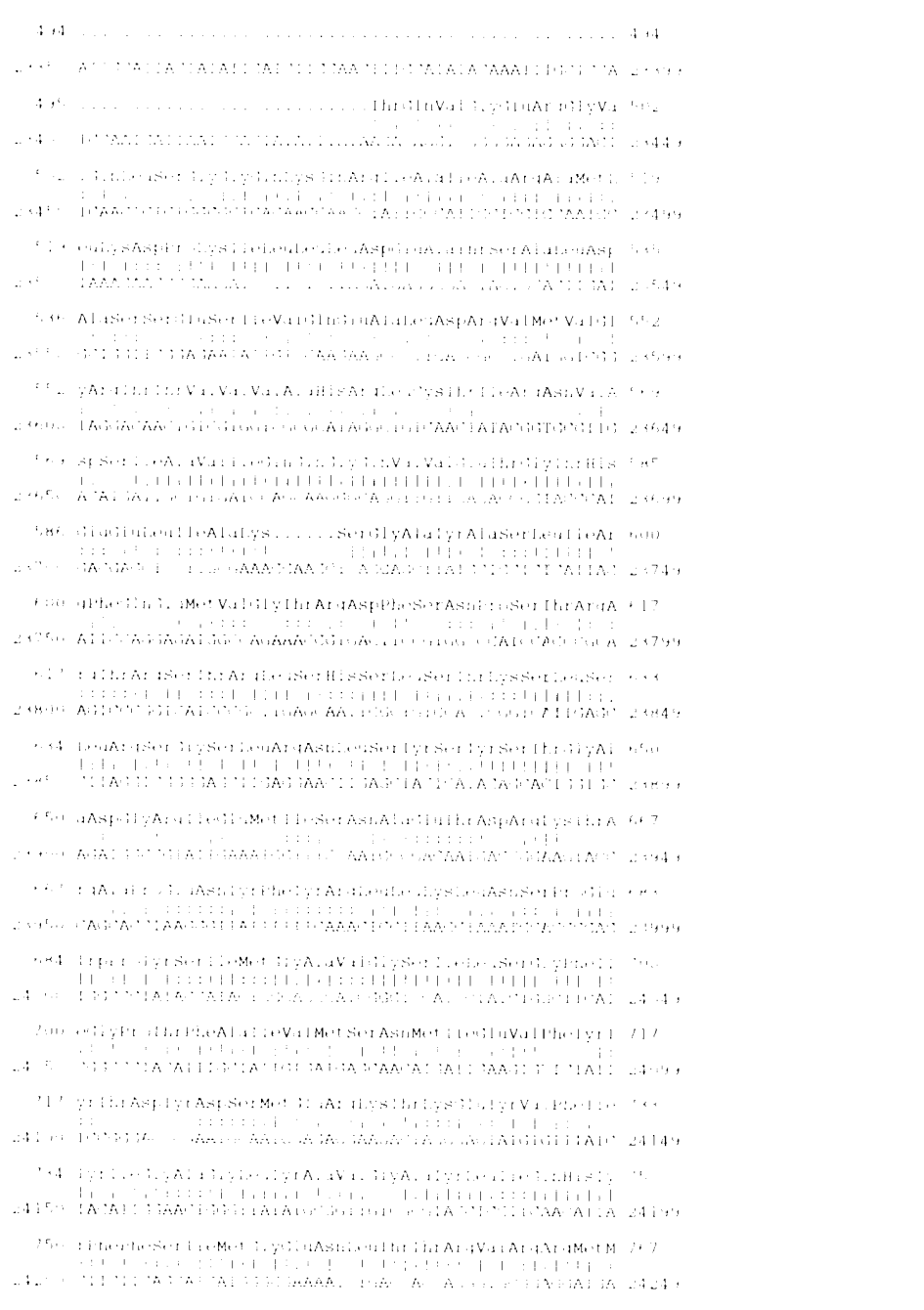


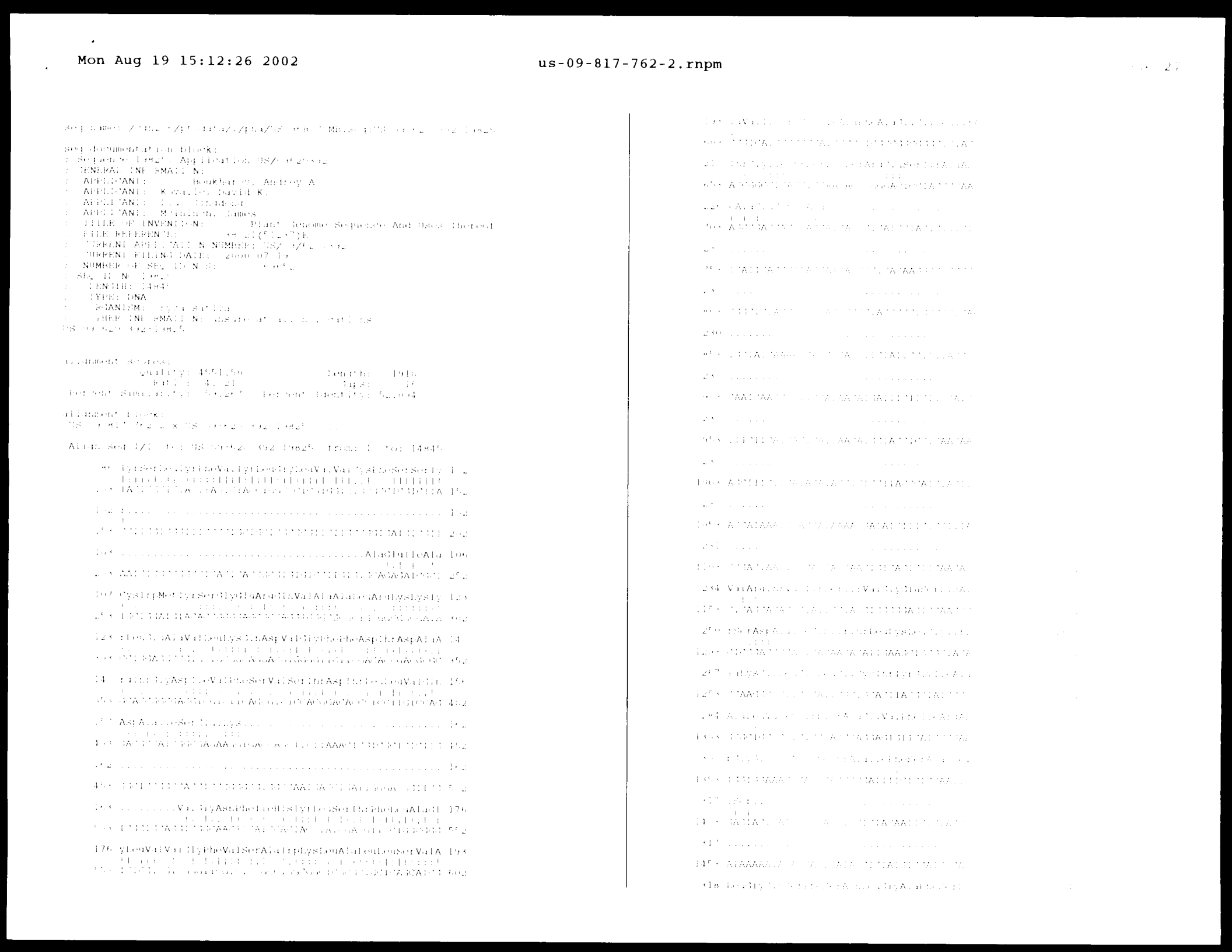


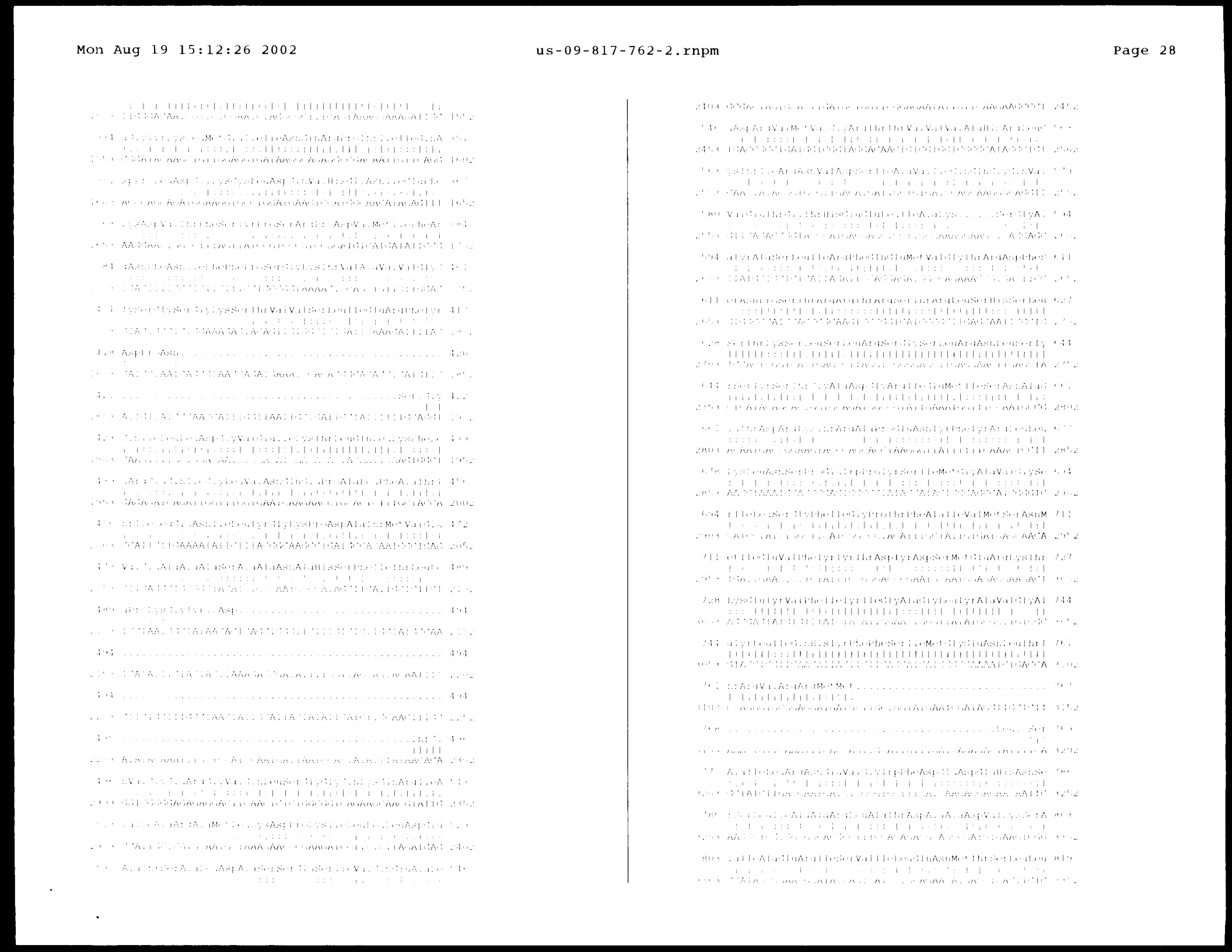


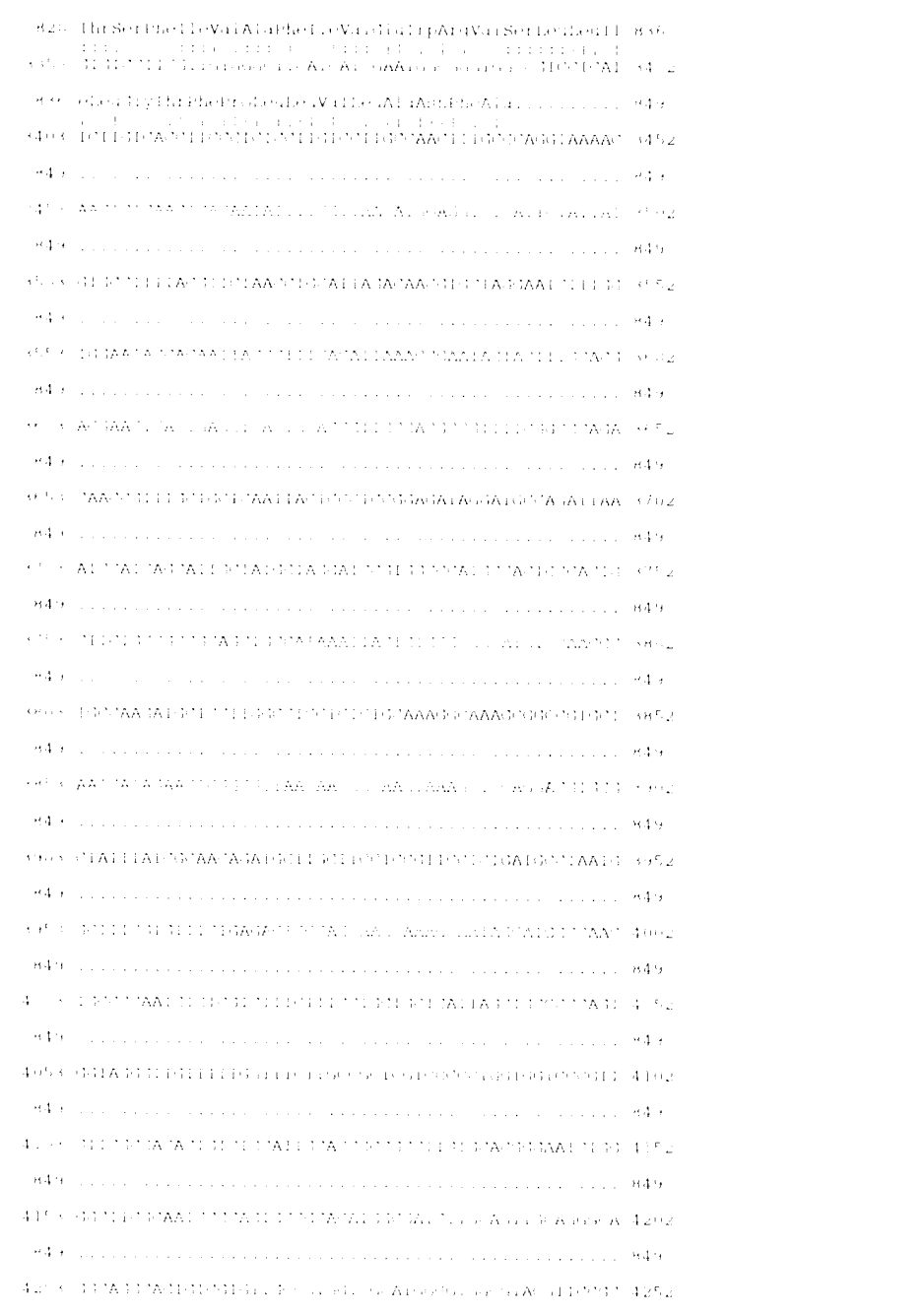


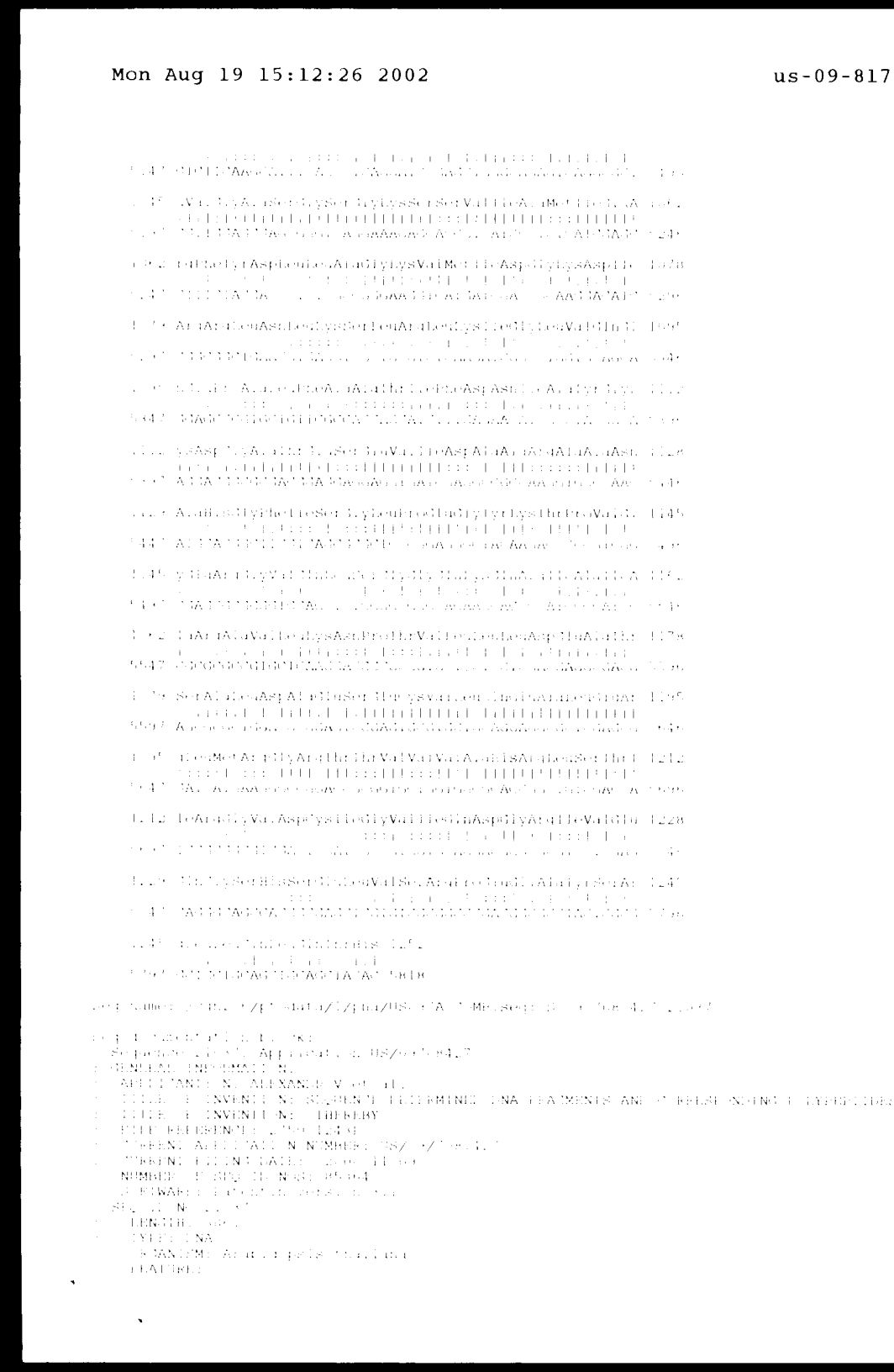


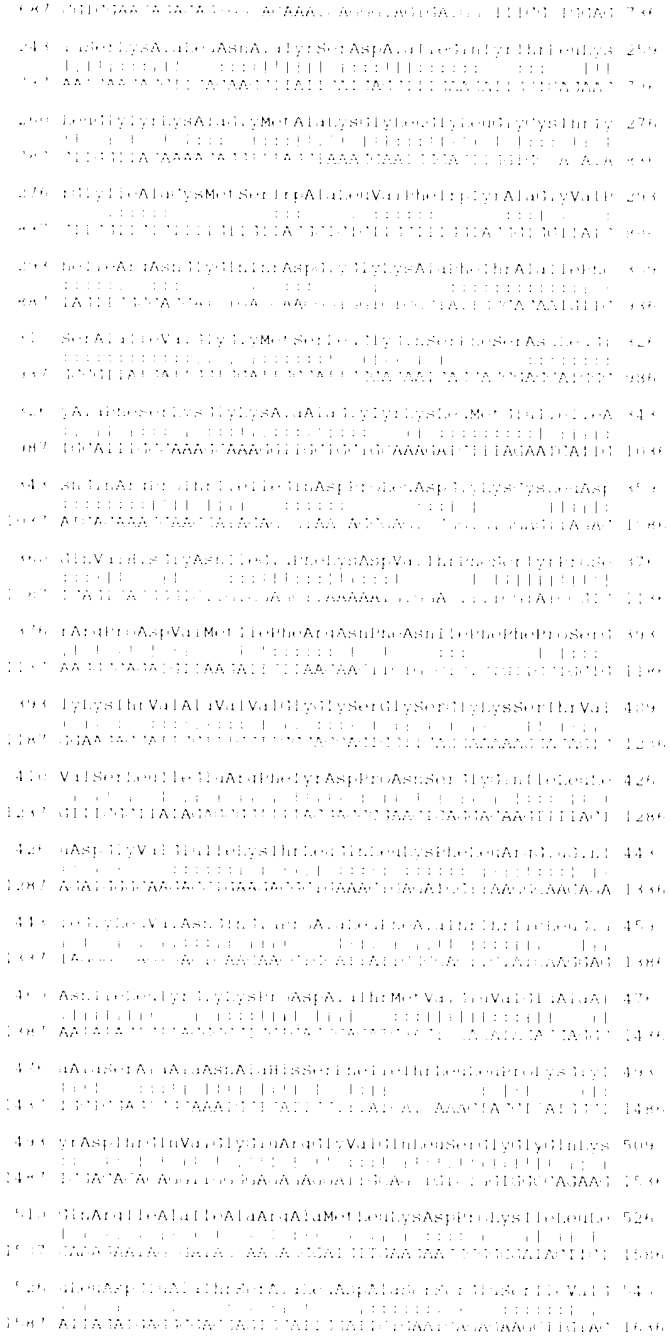


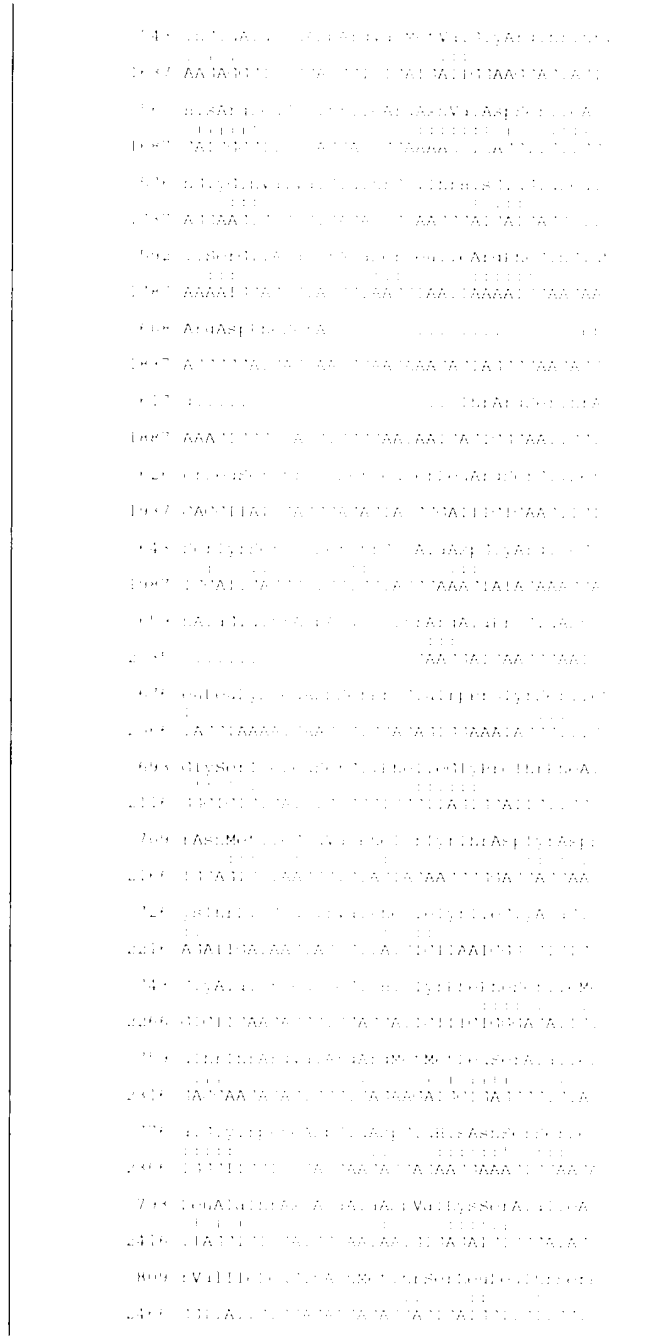


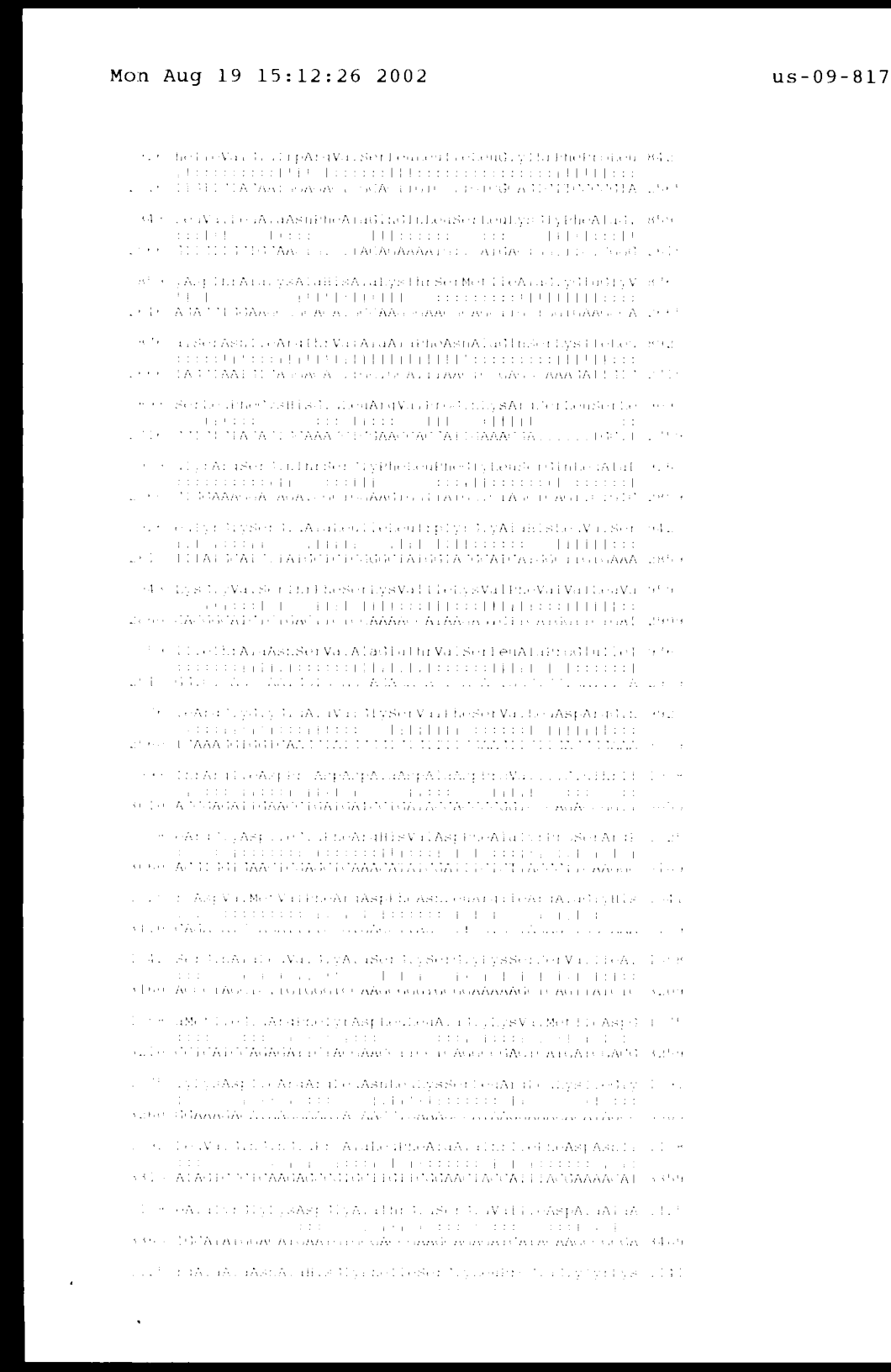


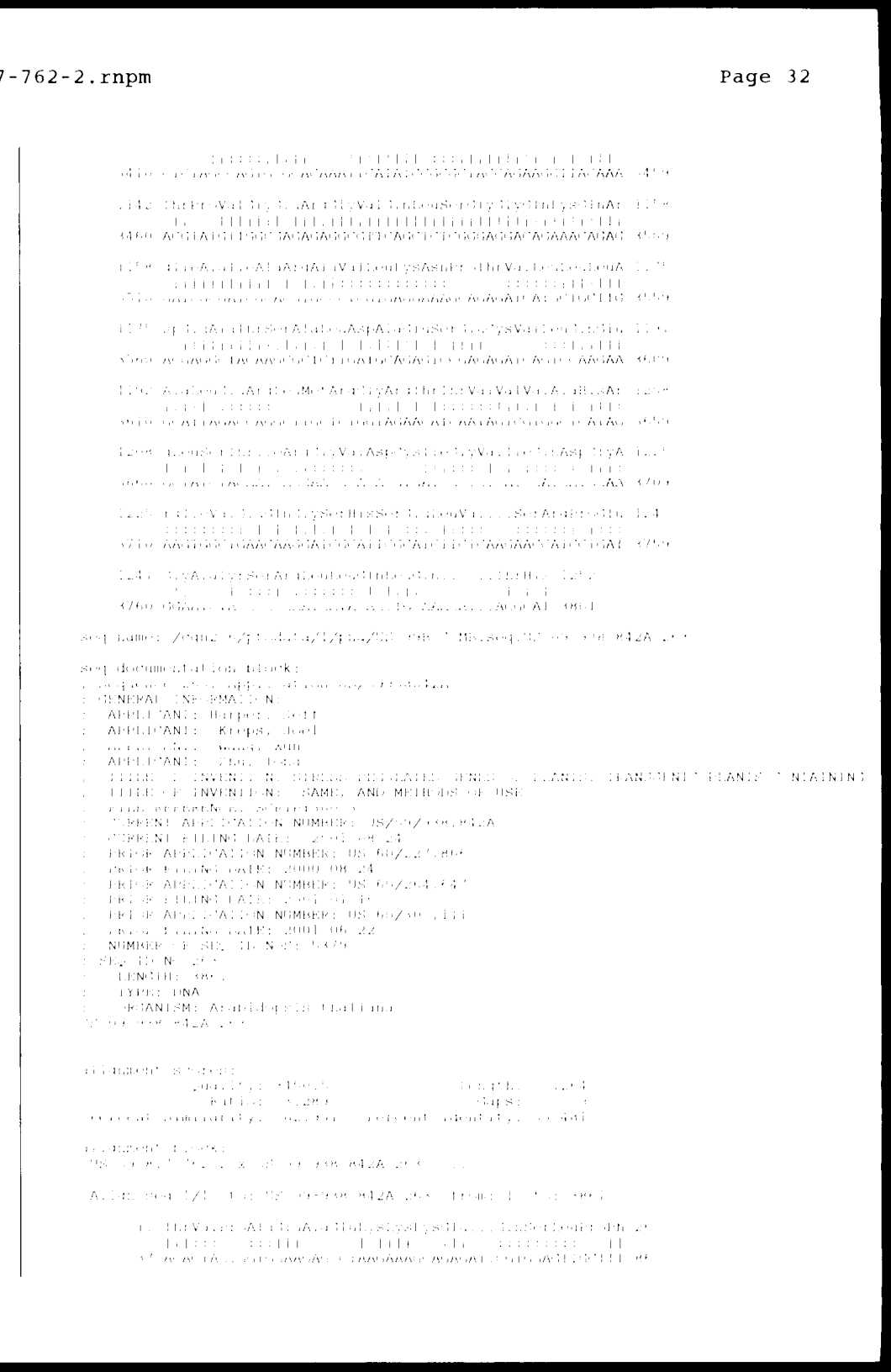


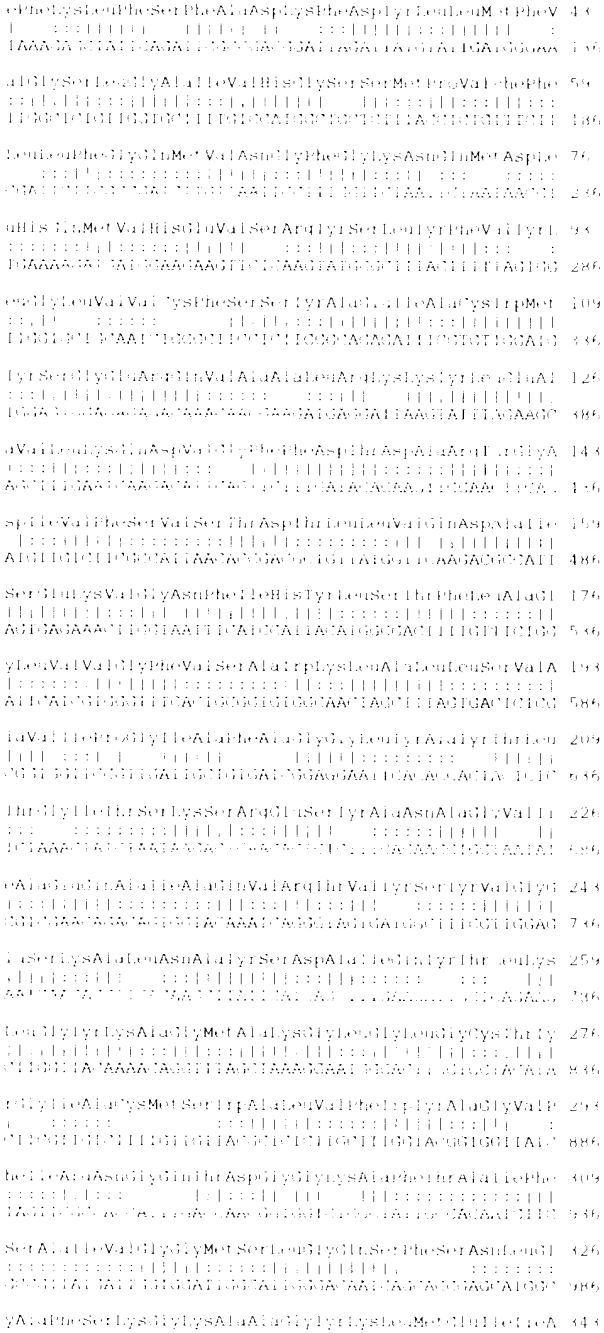


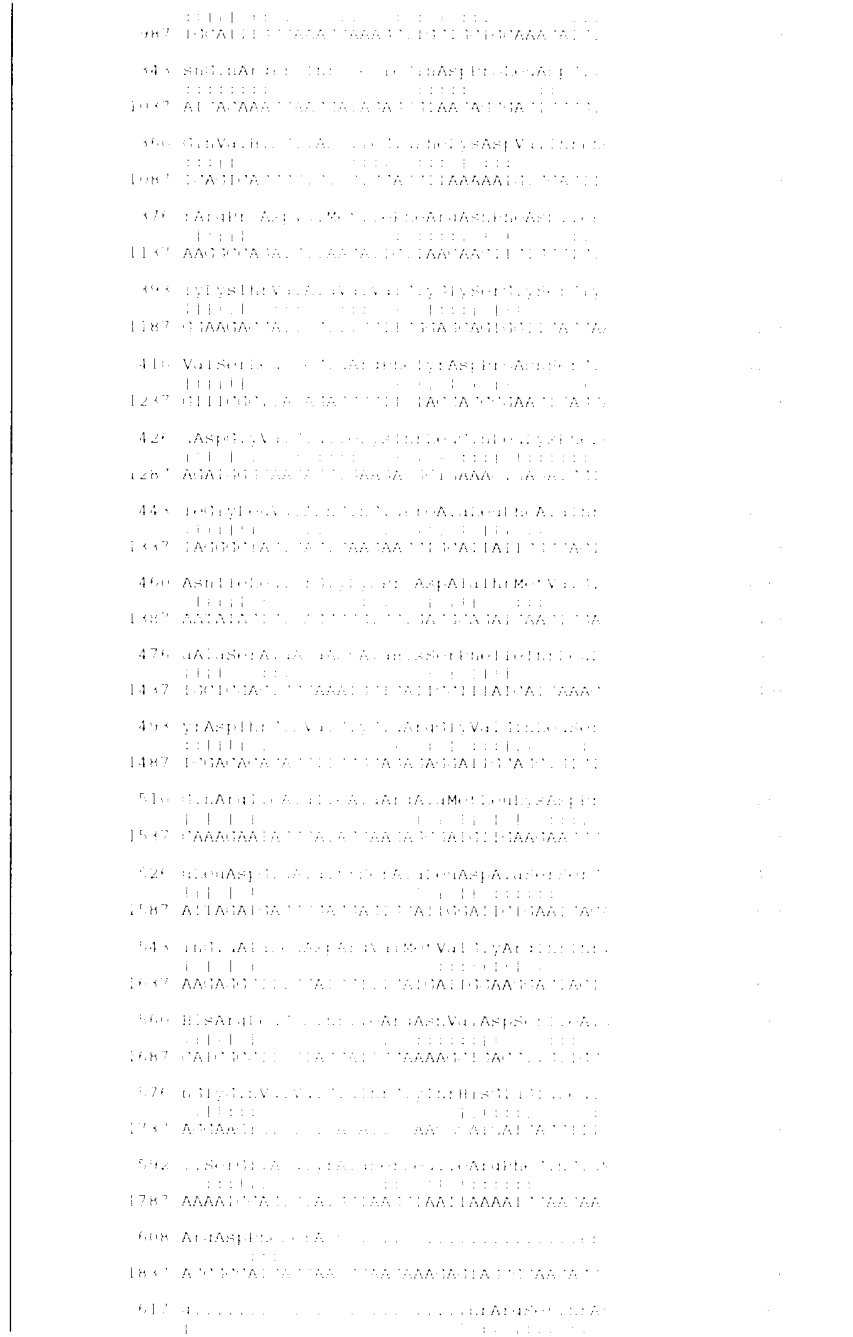




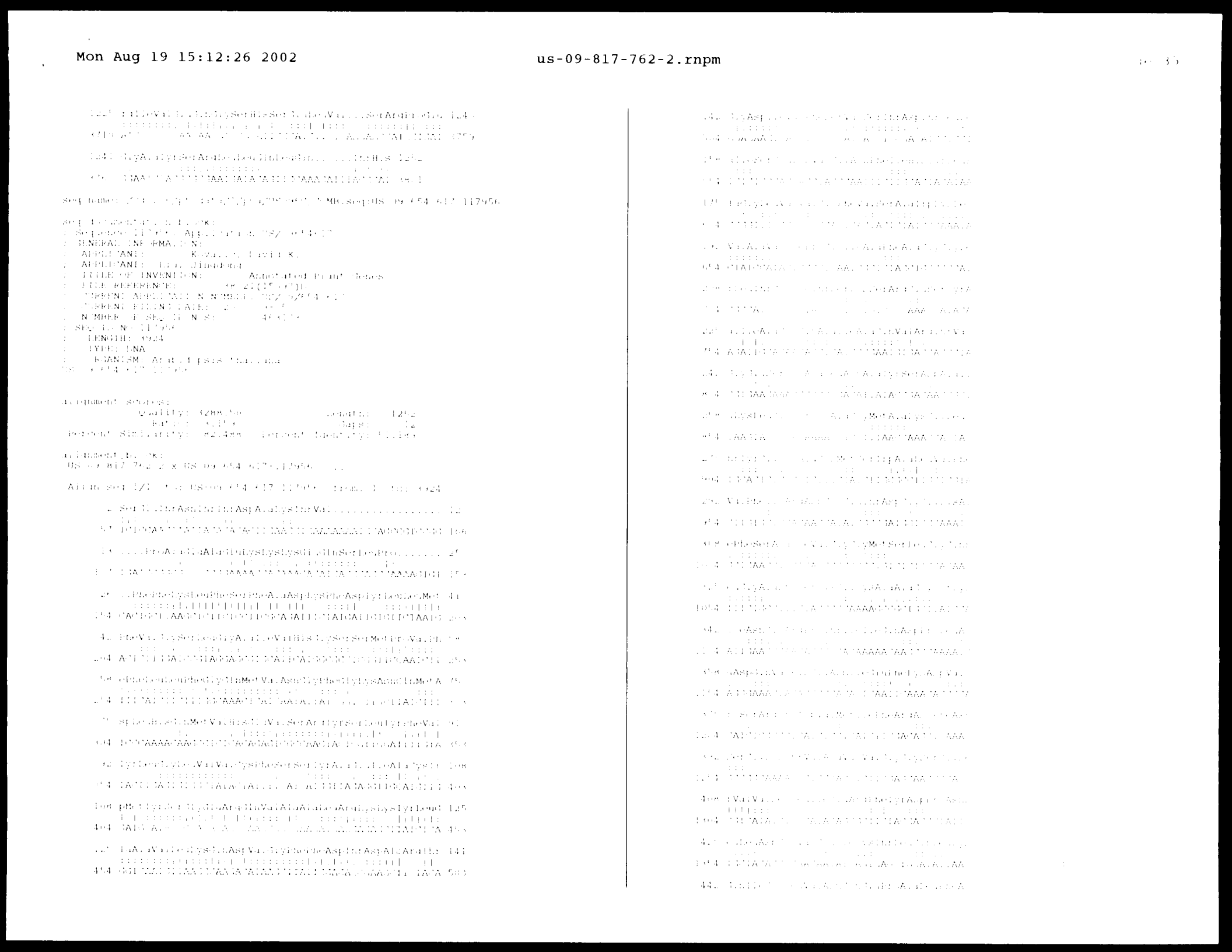


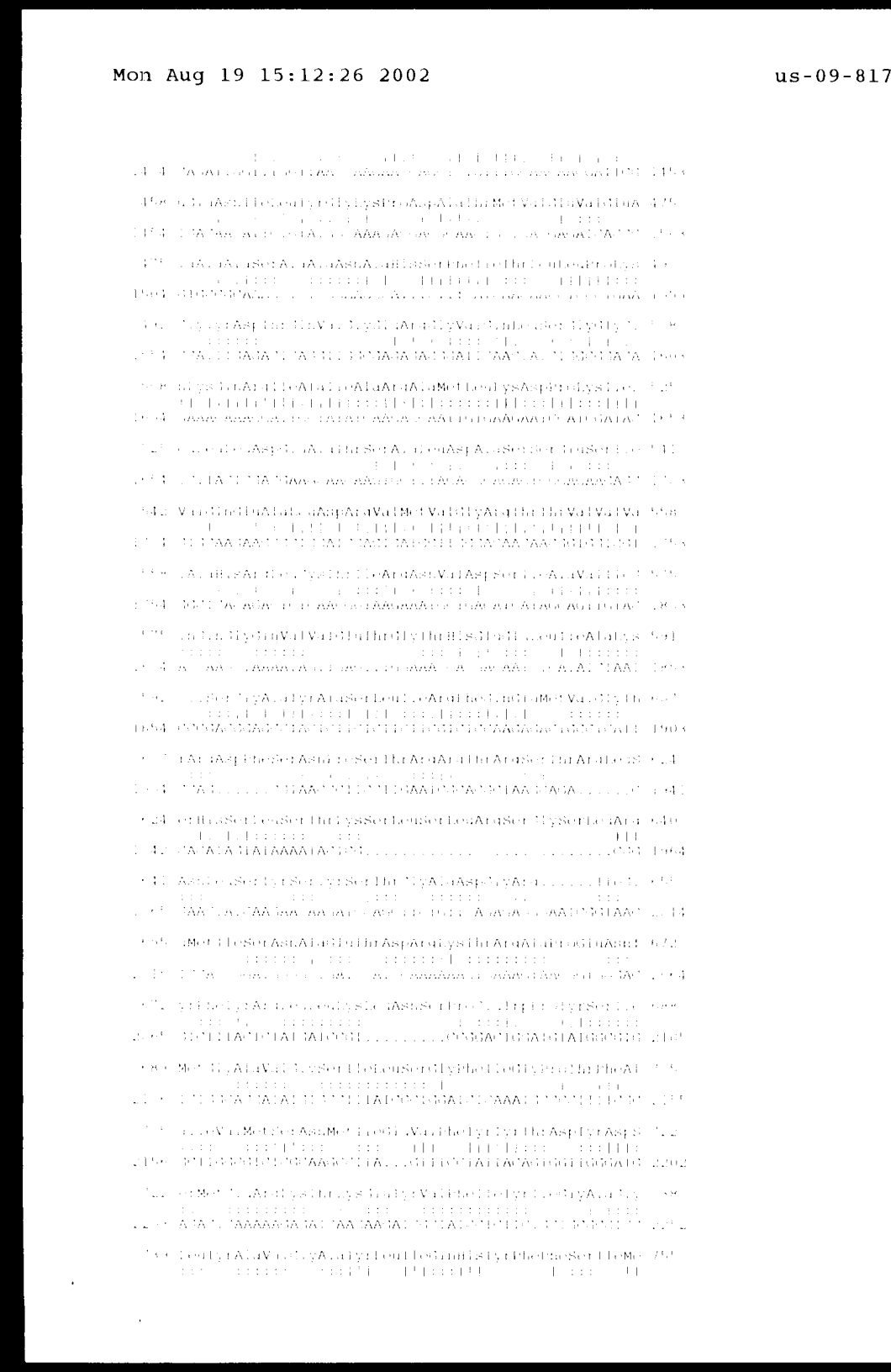


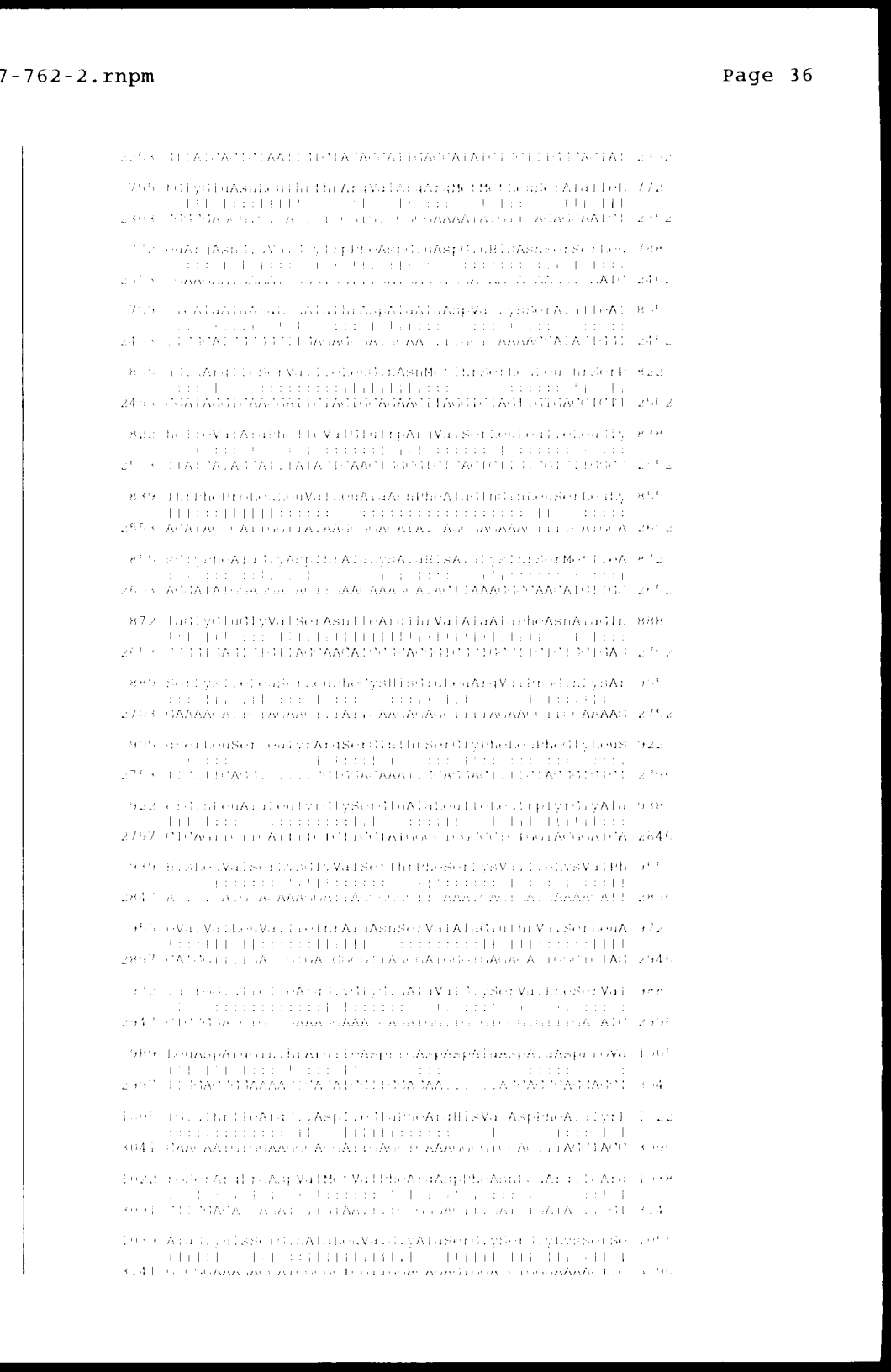


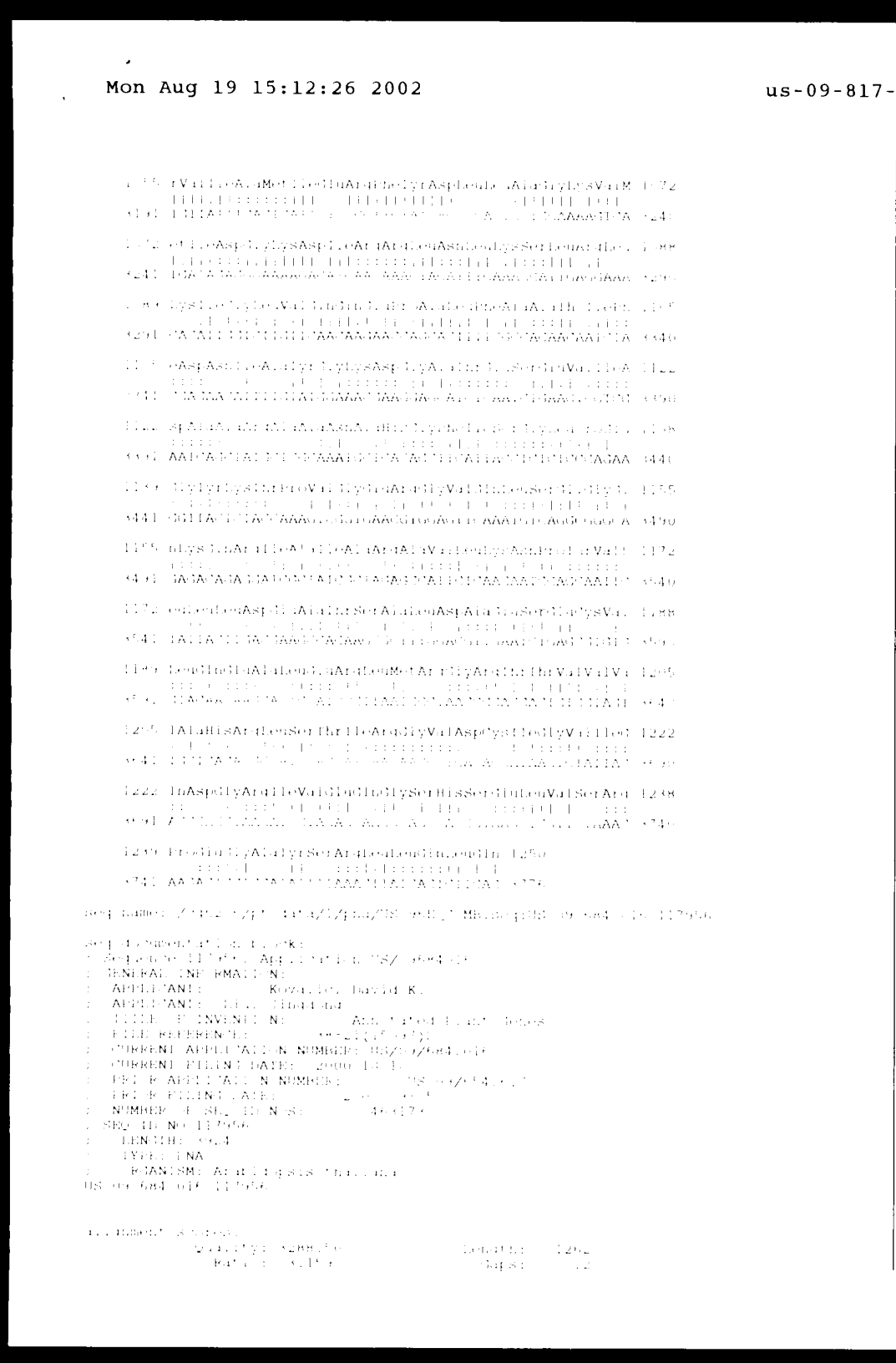


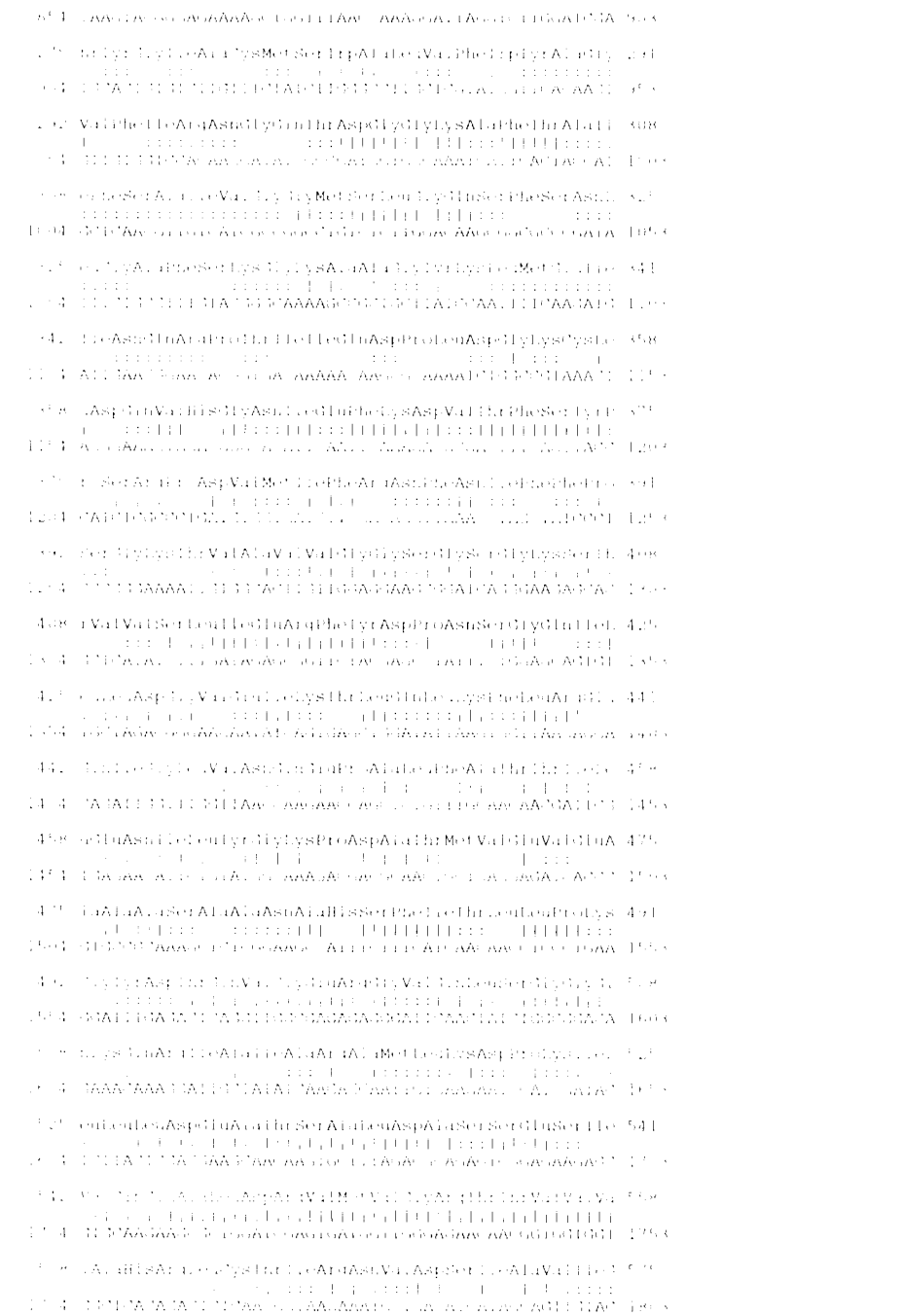


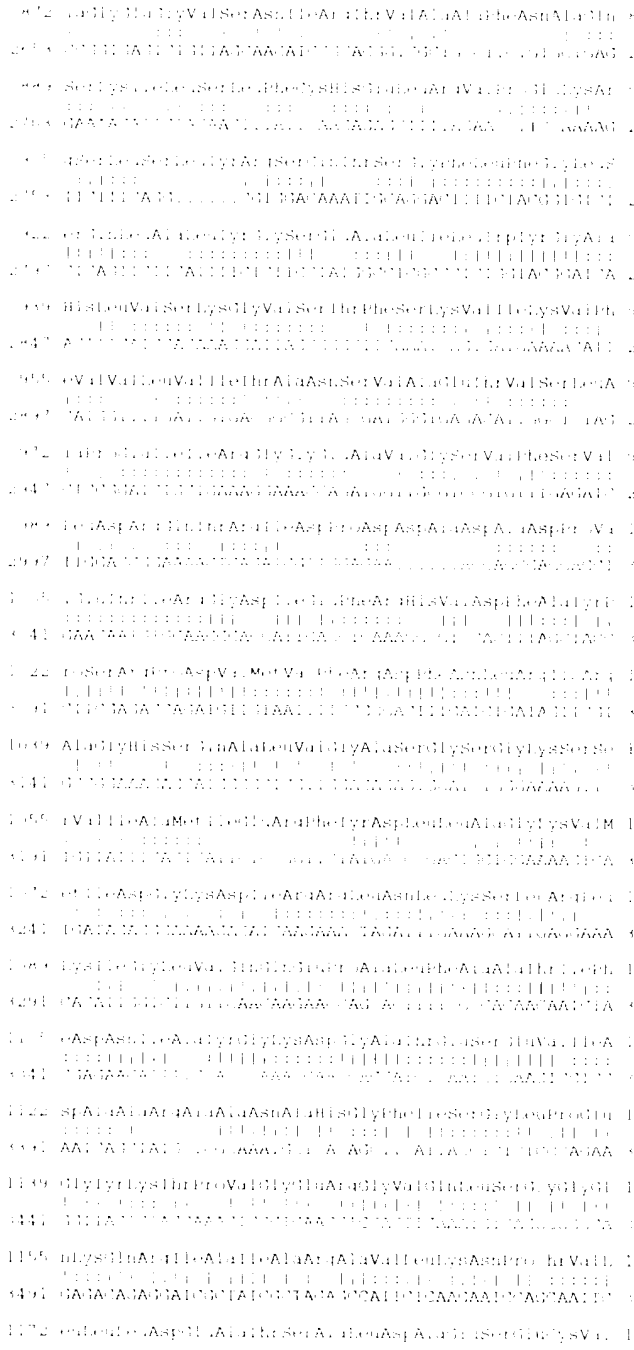


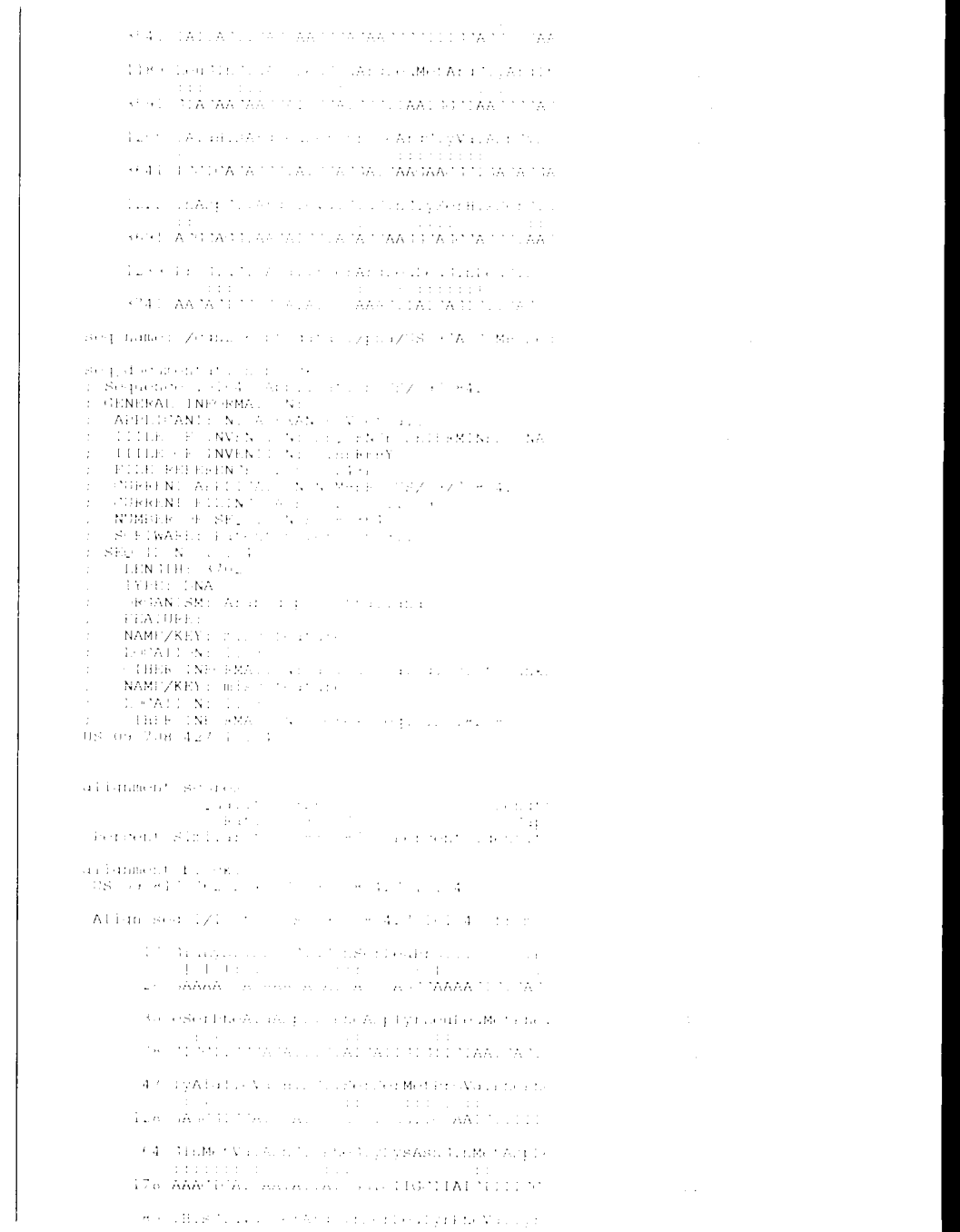


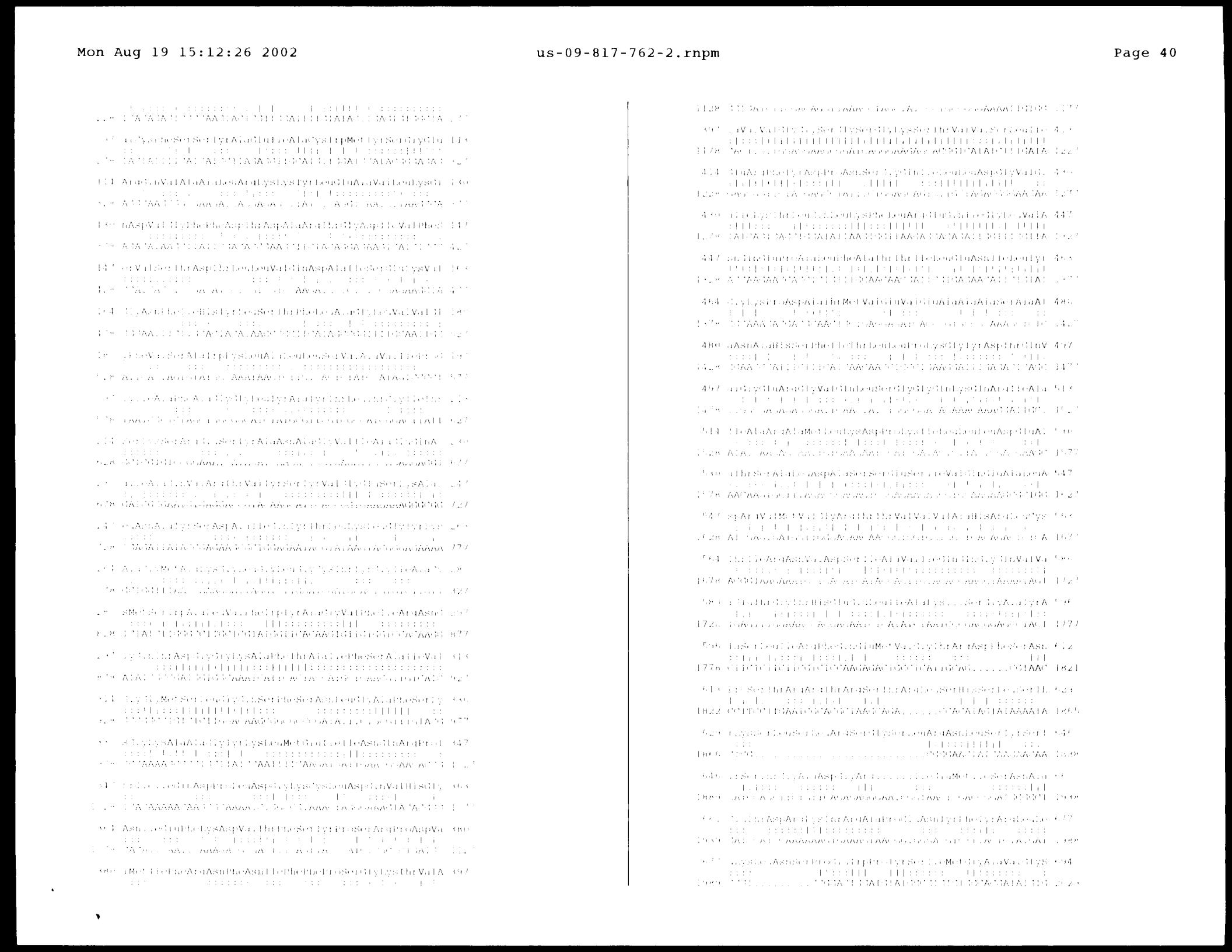


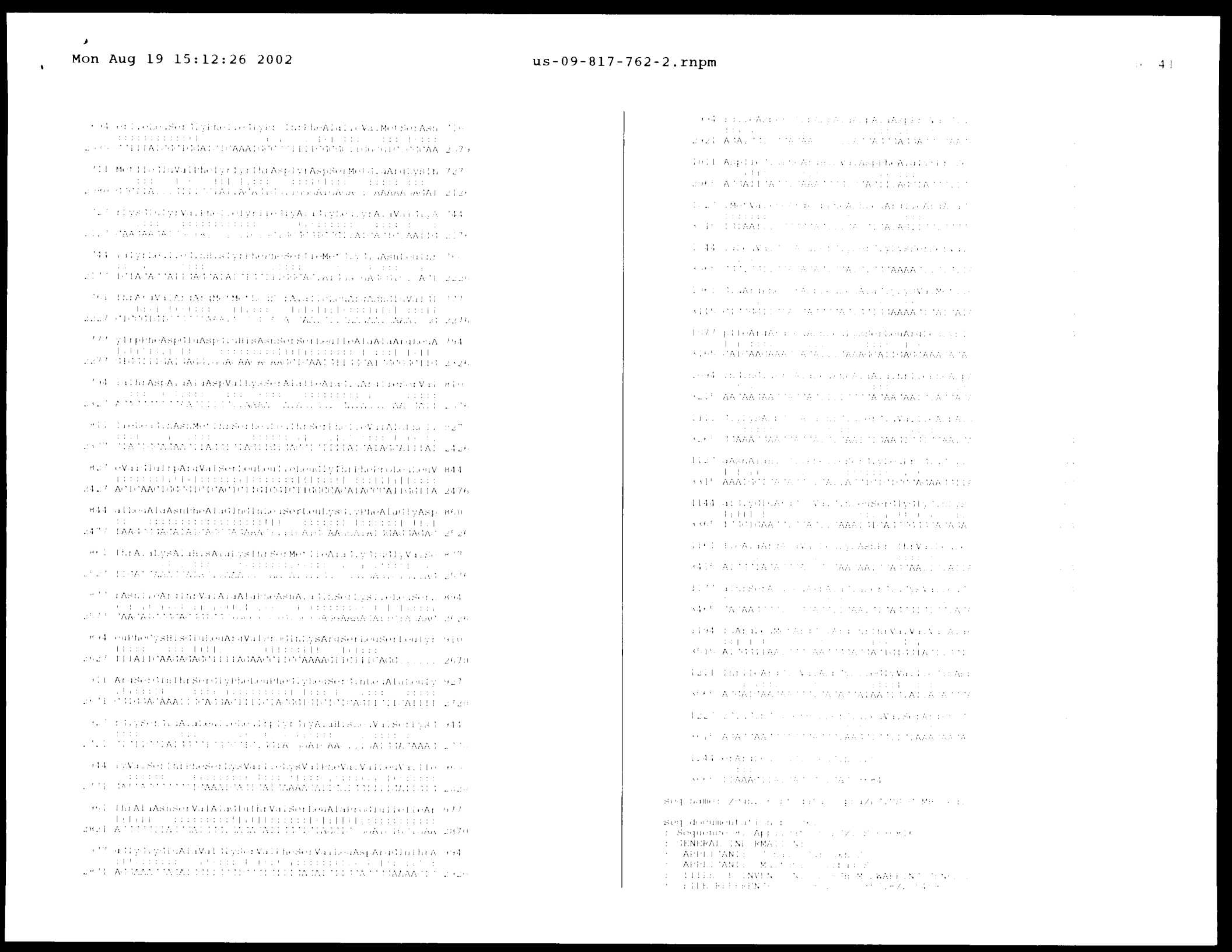


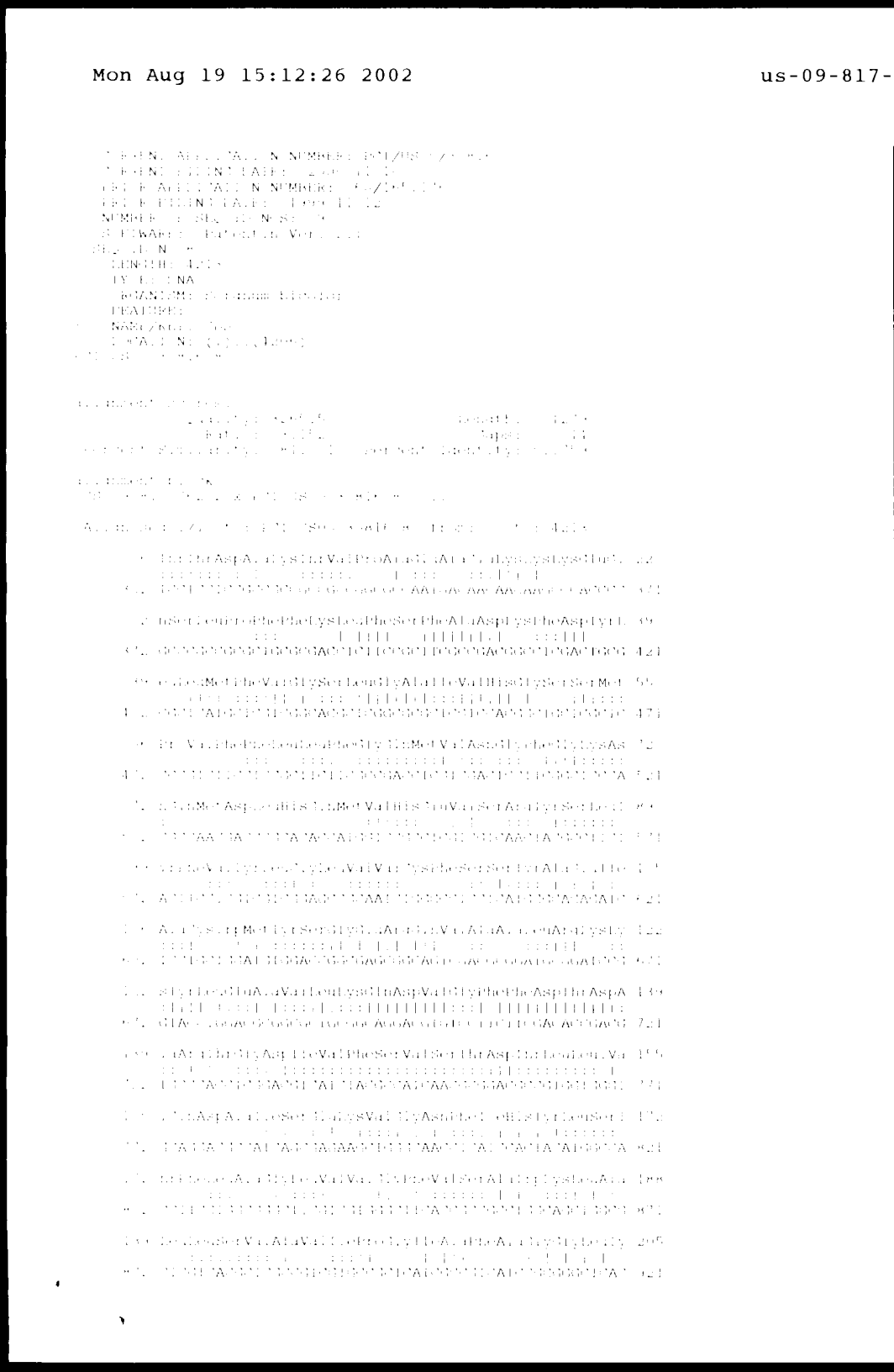


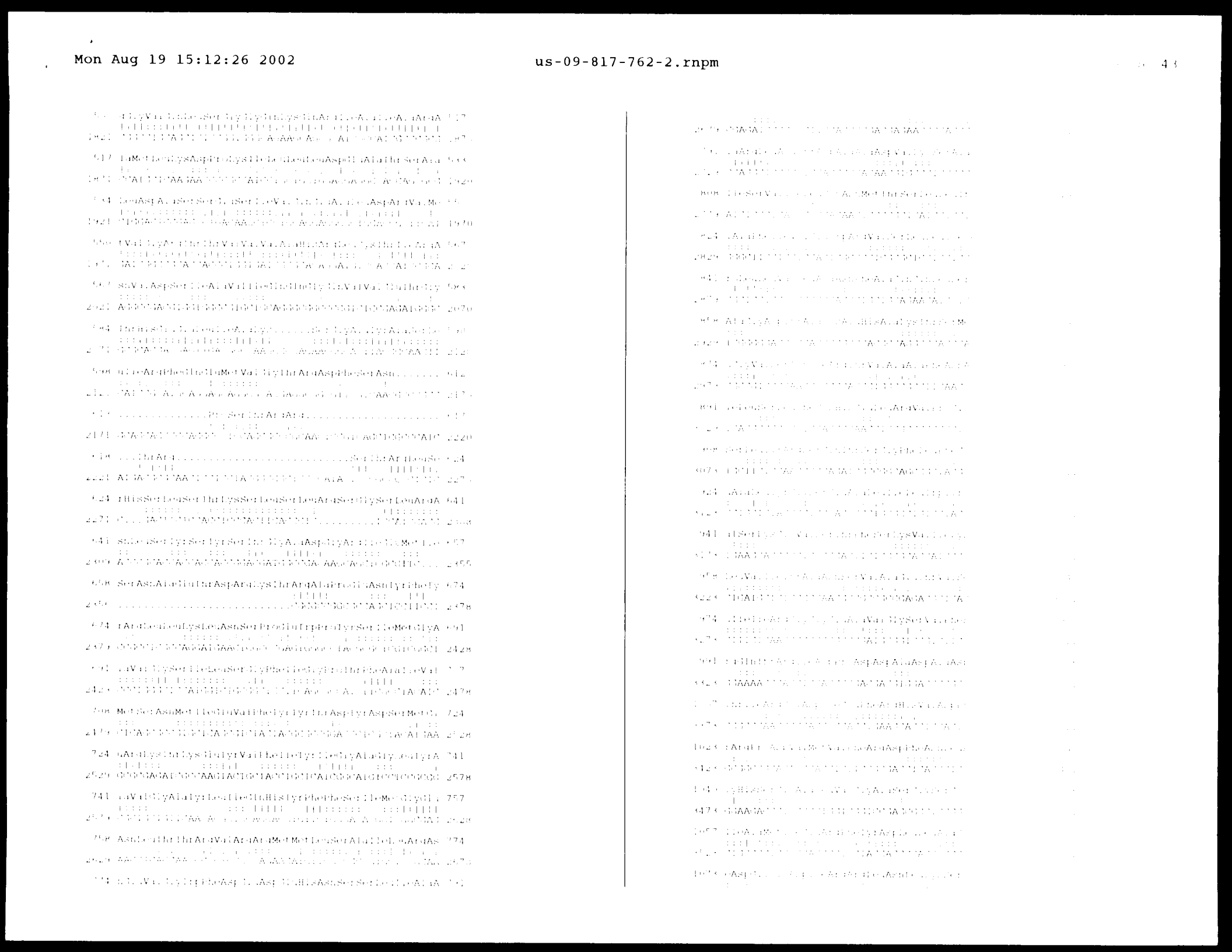


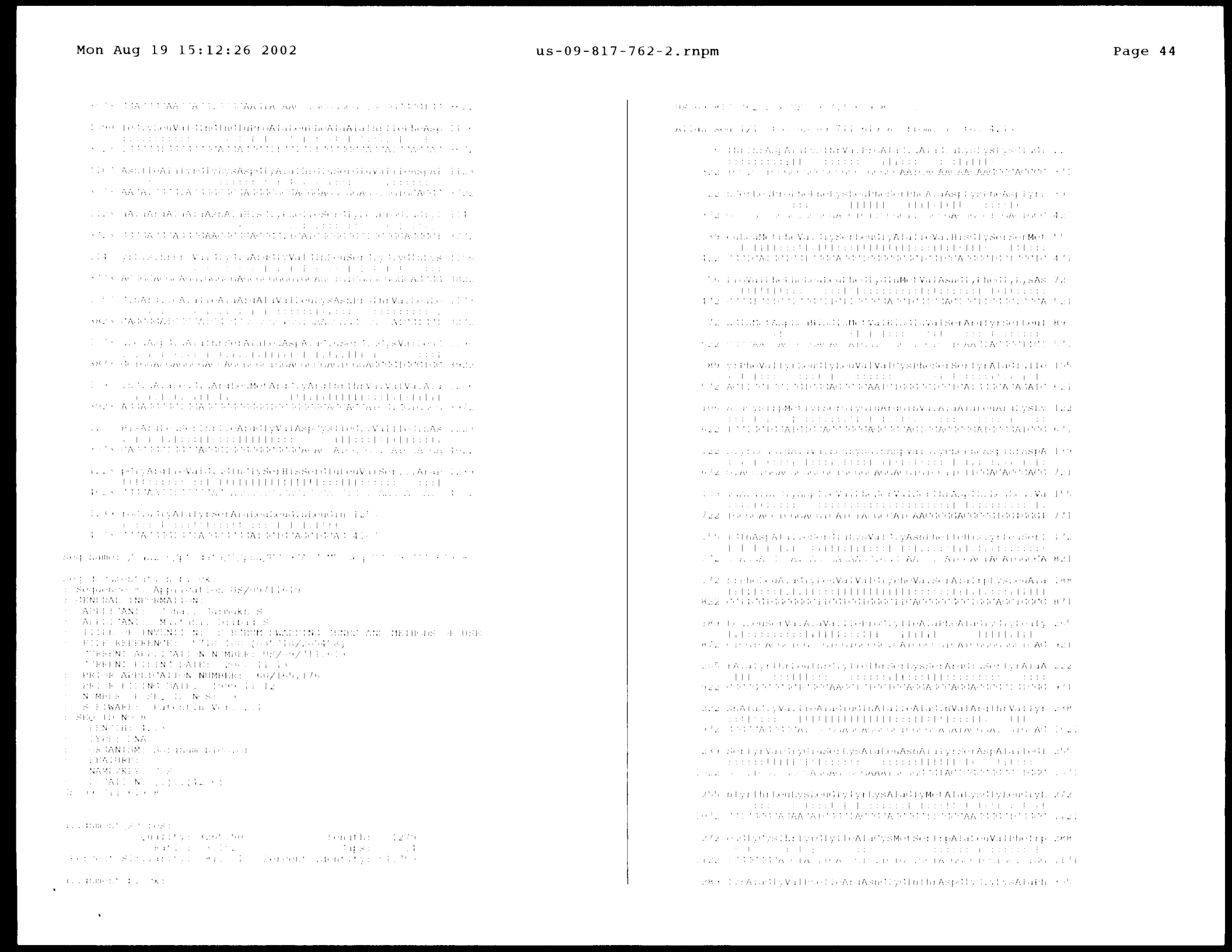


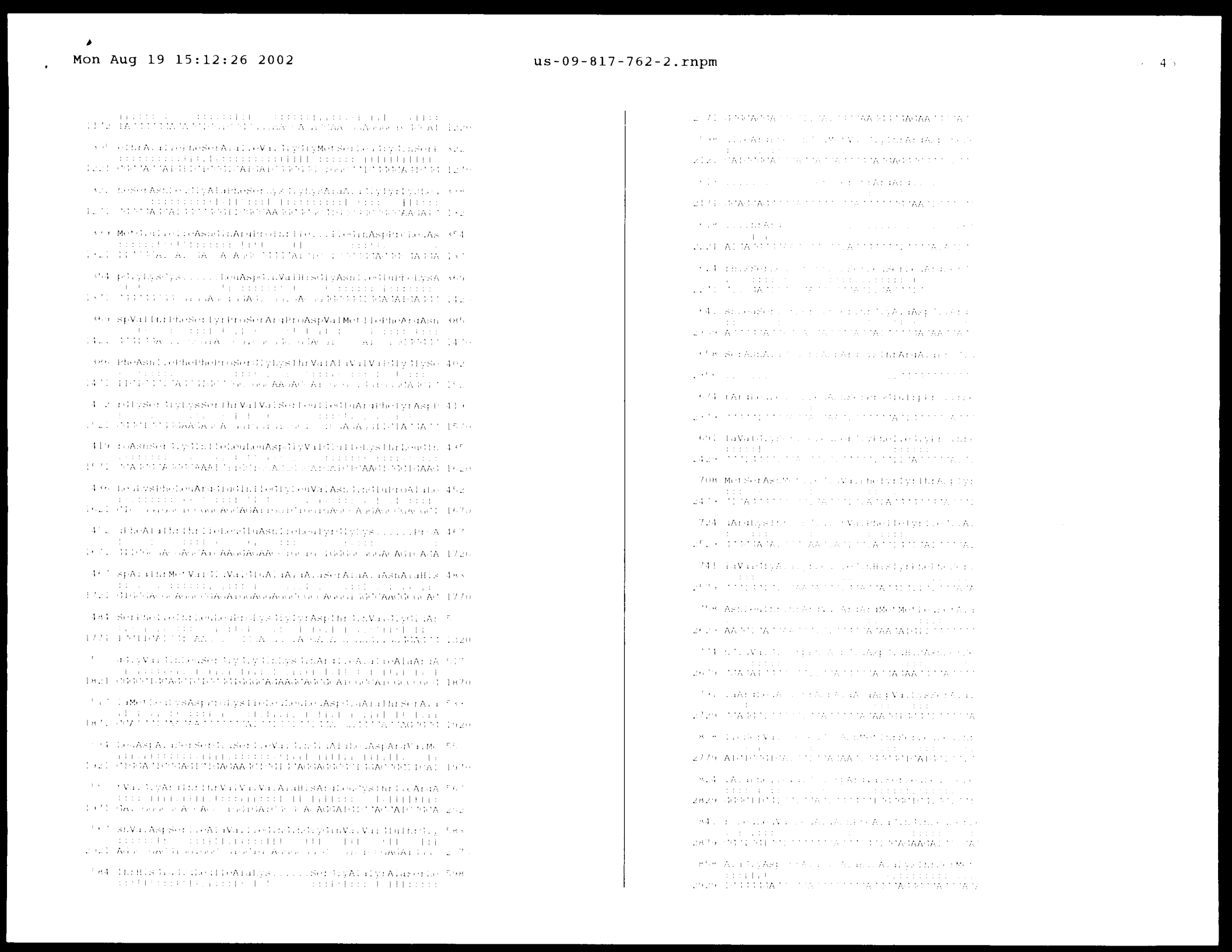


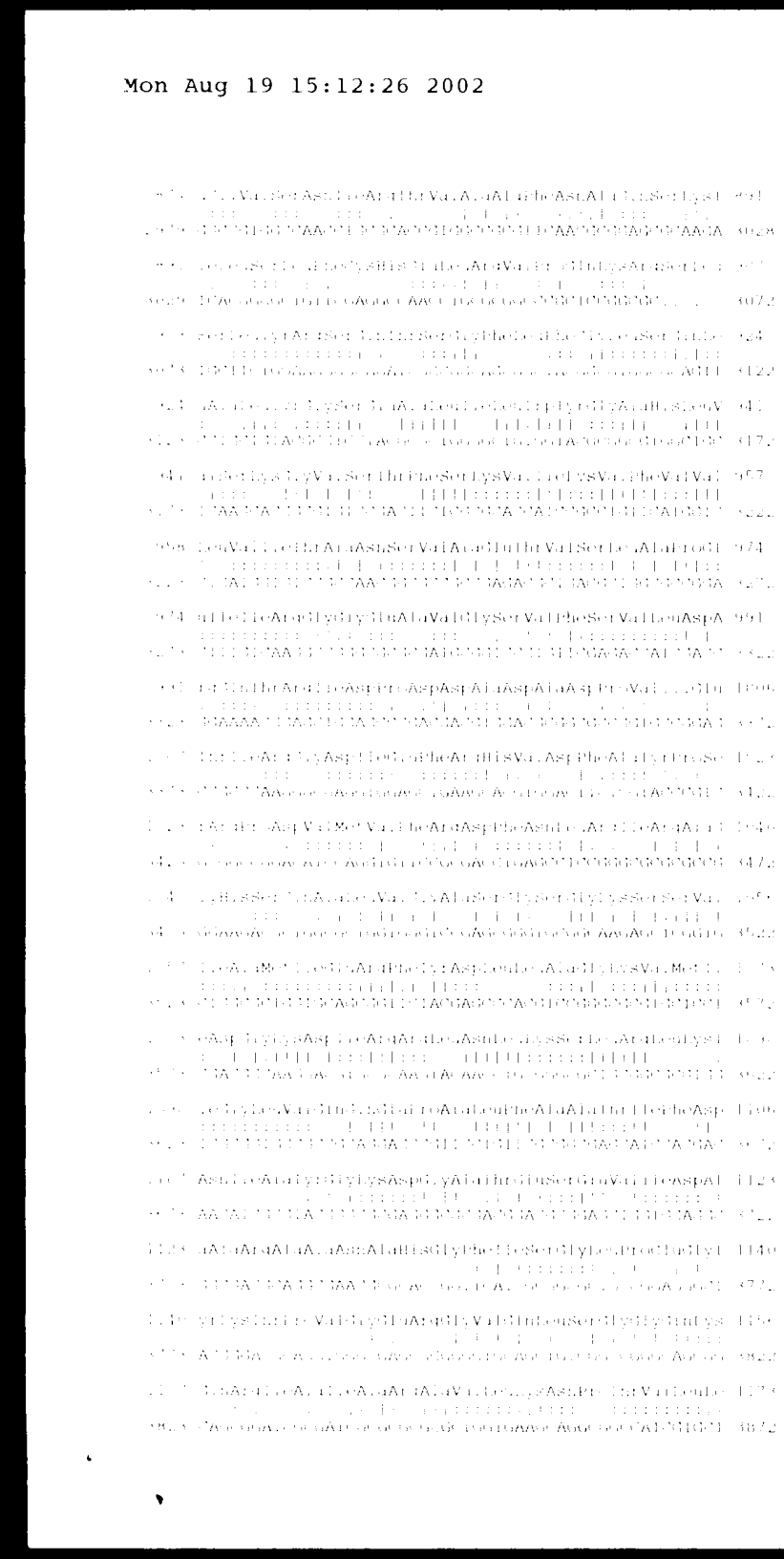


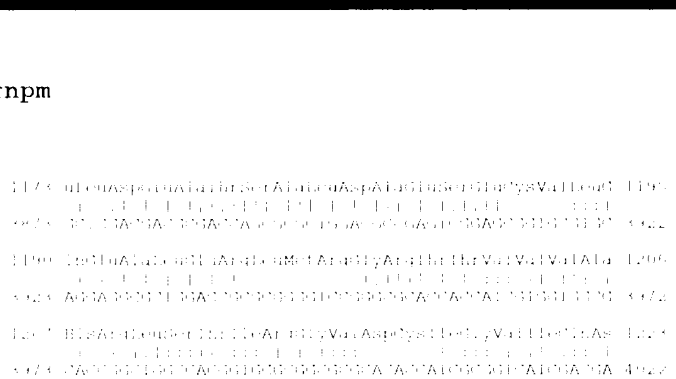


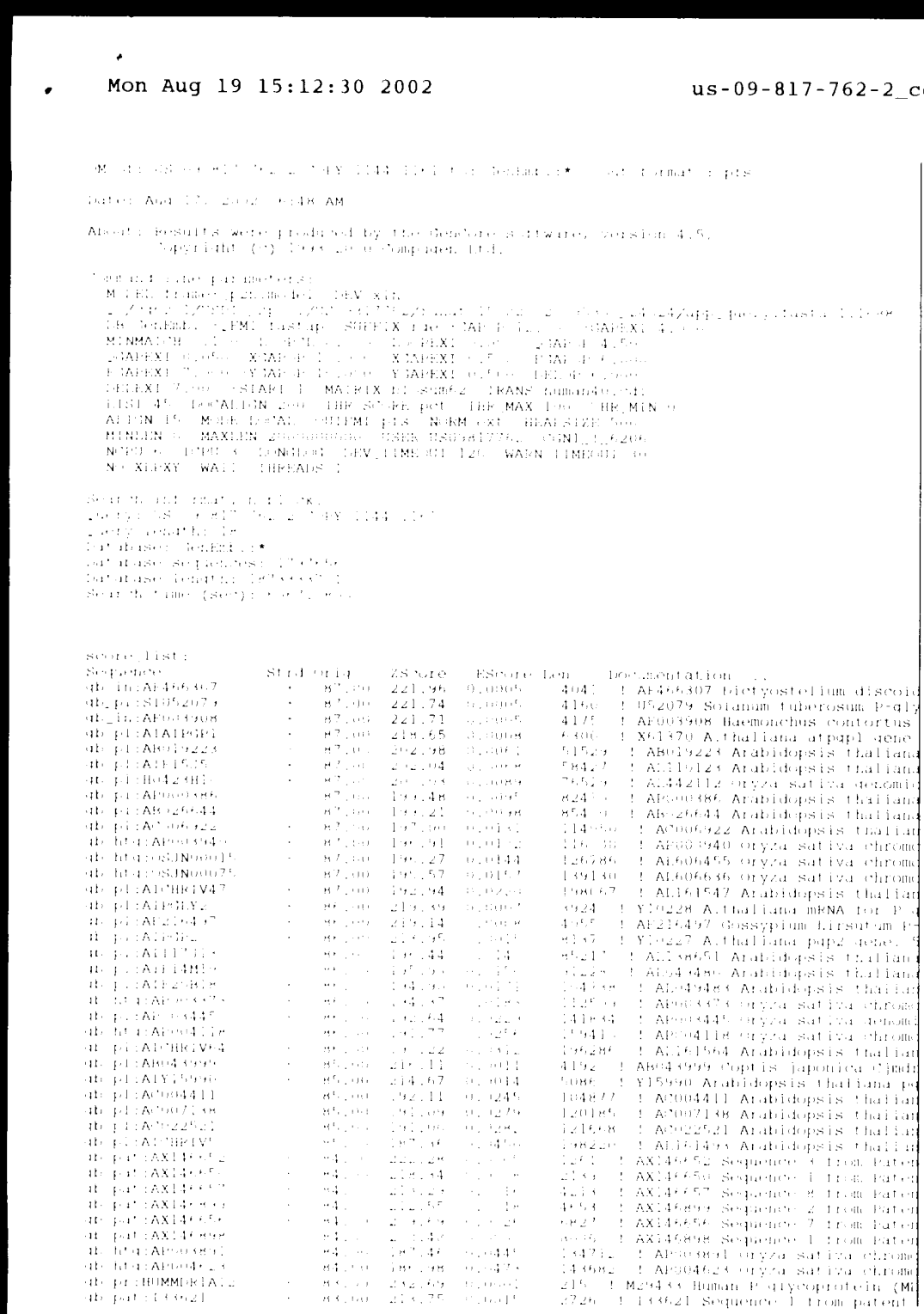


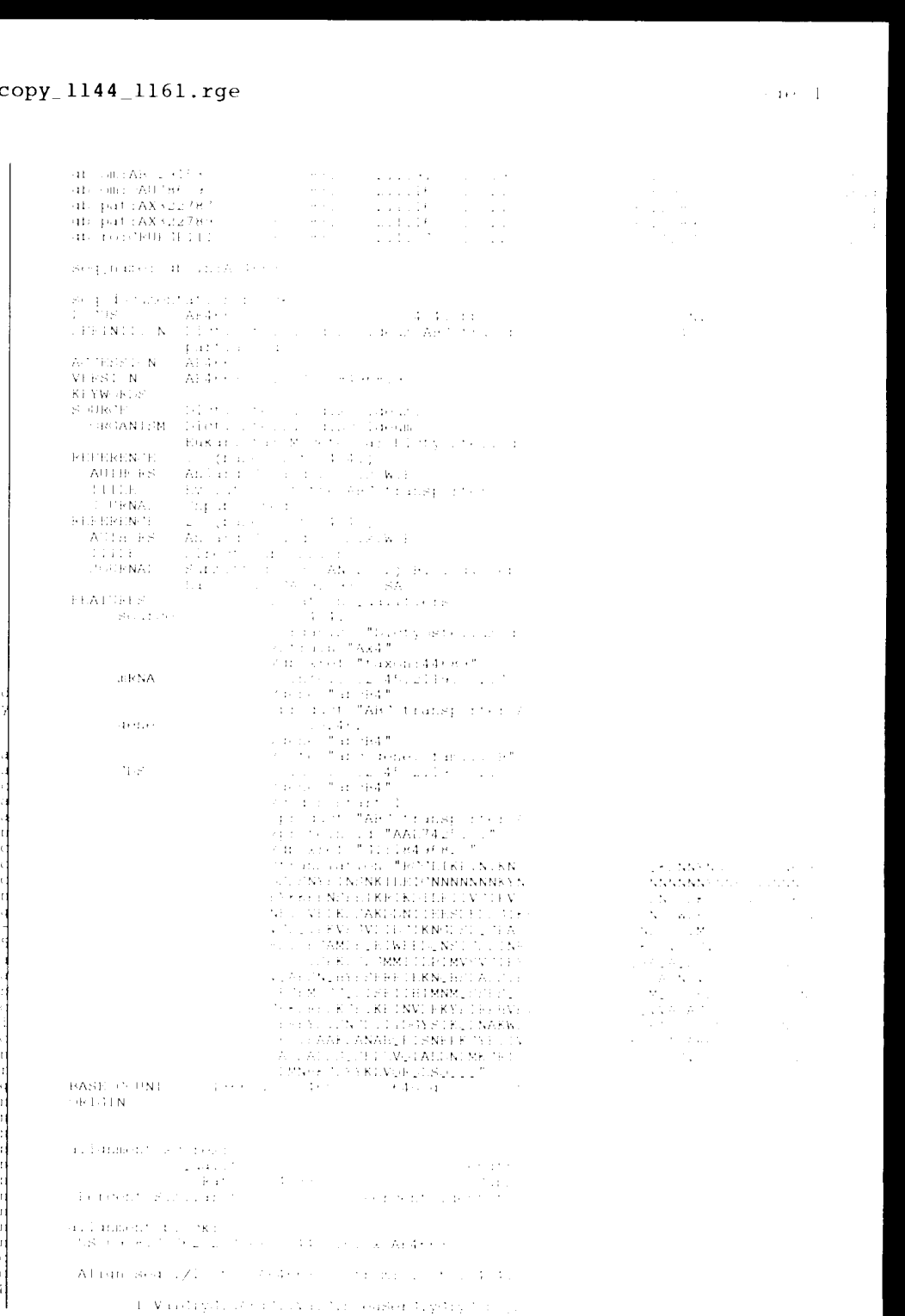


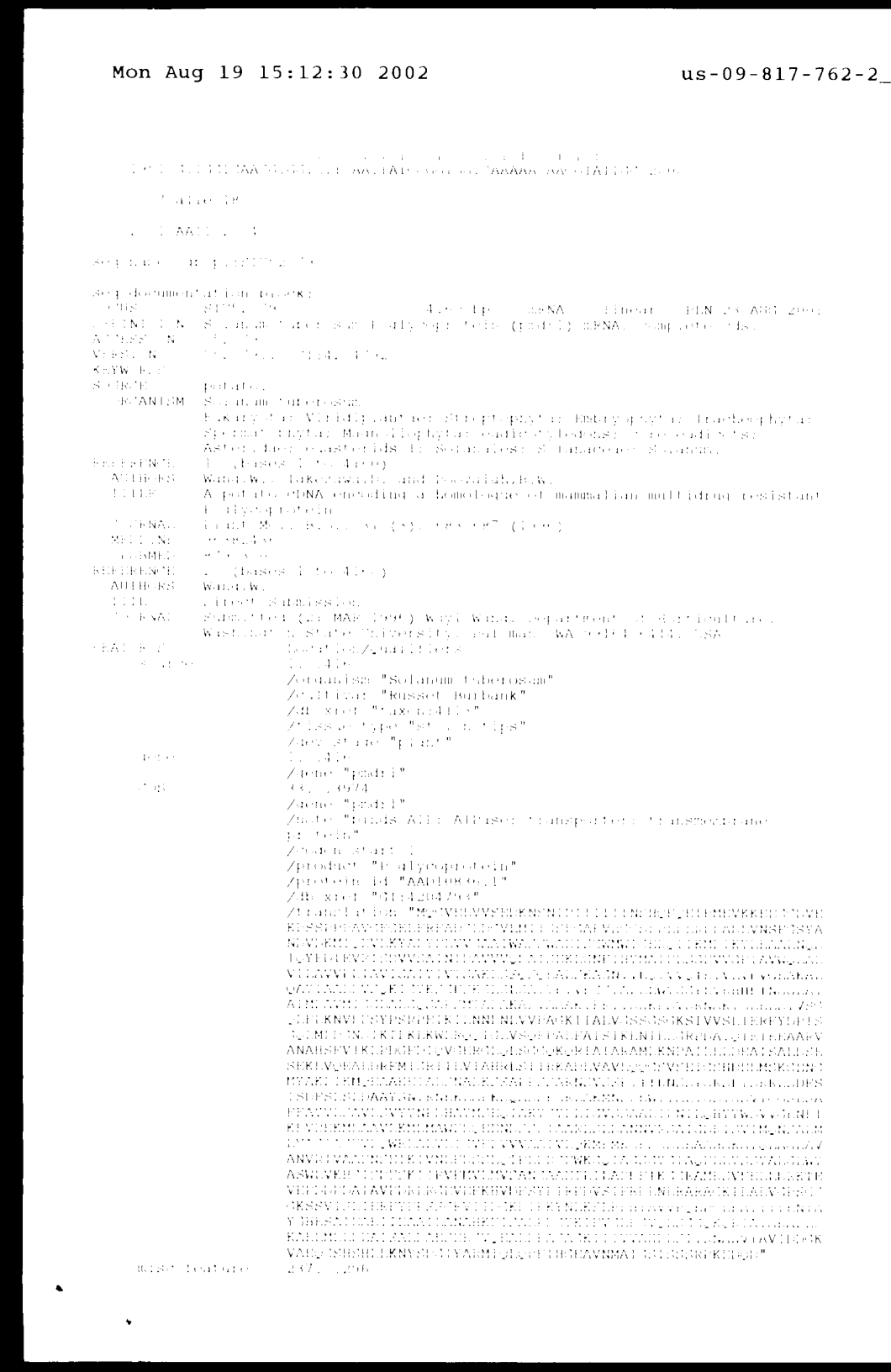


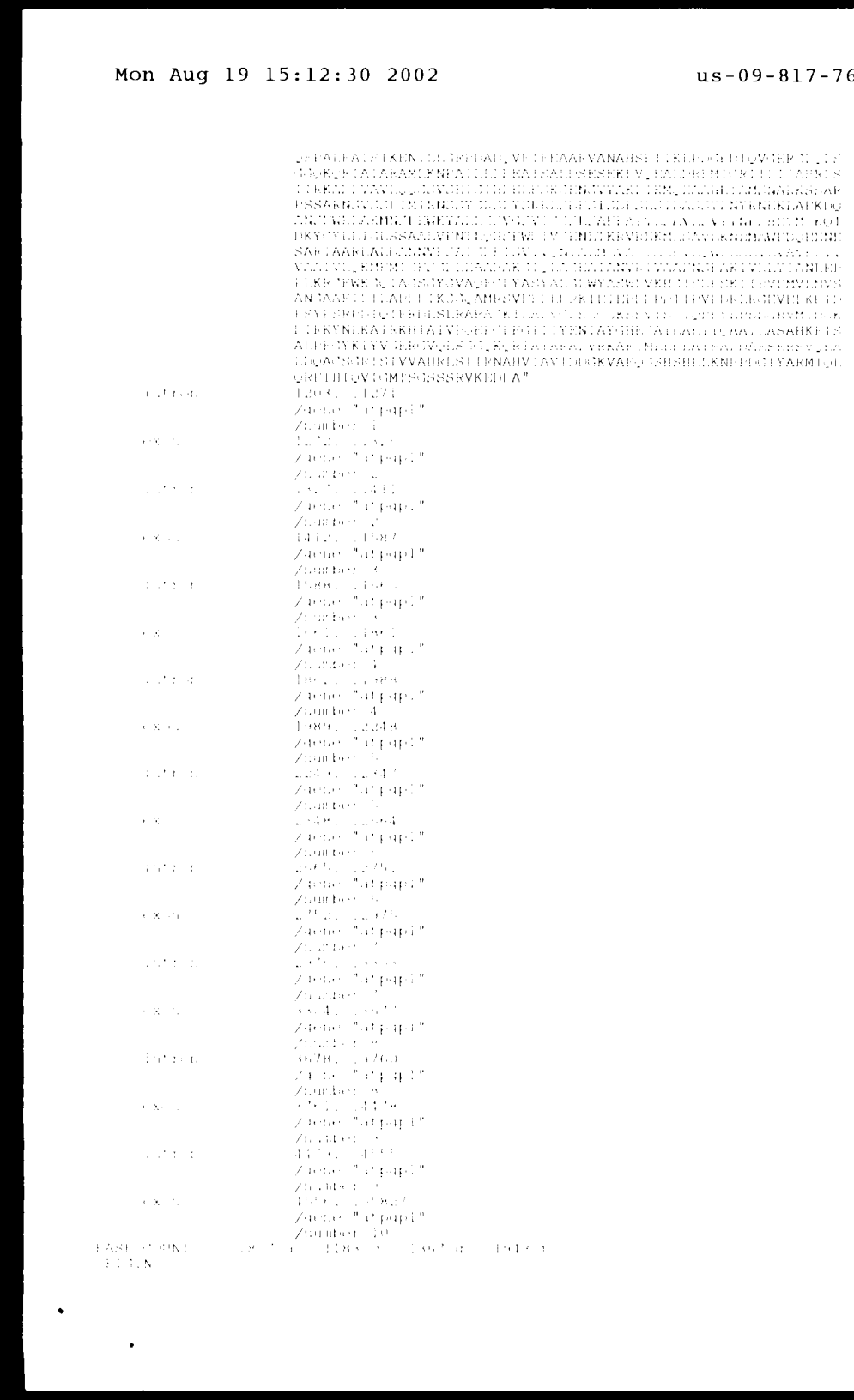


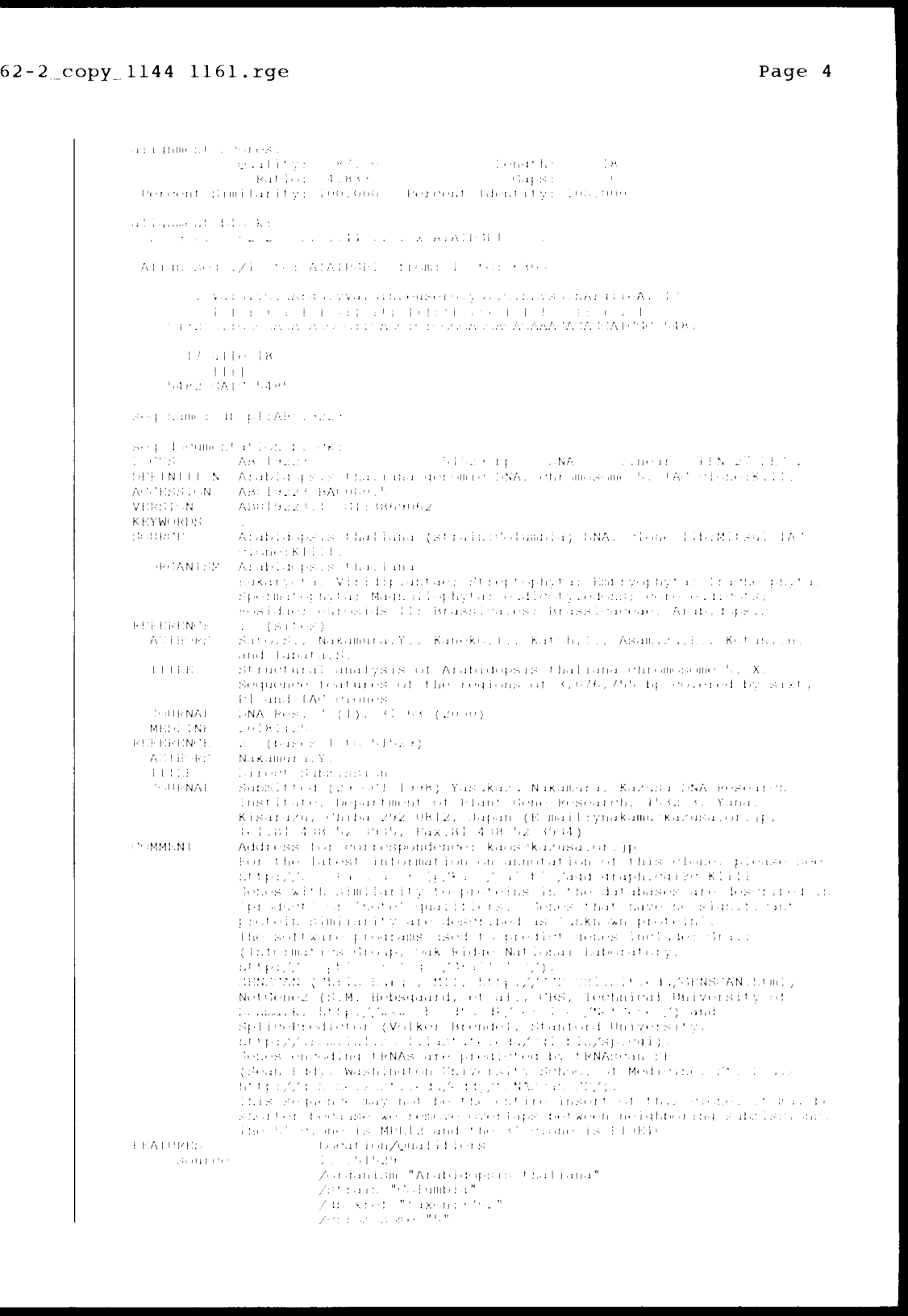


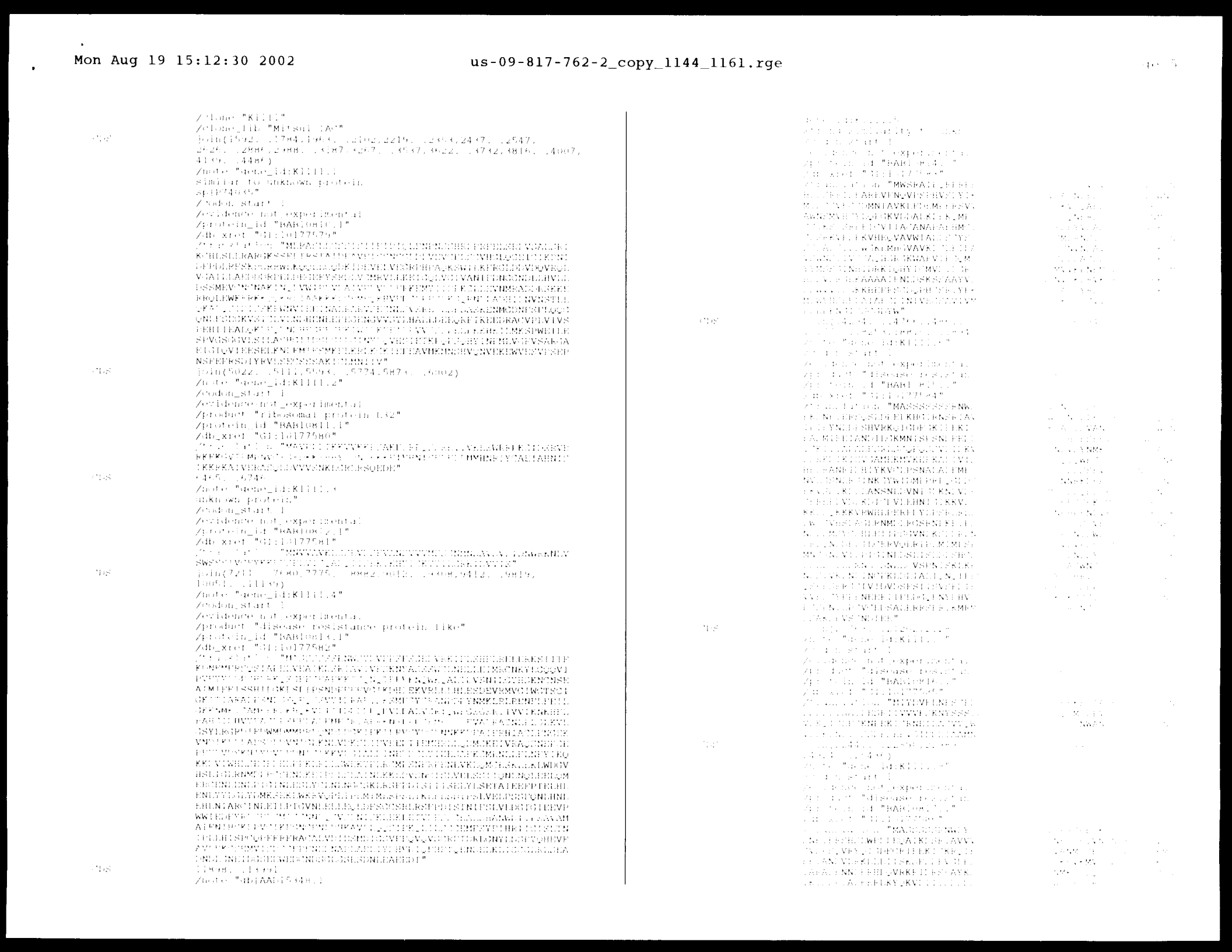












```
                    [illegible]
                    KYLRGRNEEYWMD[illegible]

alignment_scores:
         Quality:   [illegible]              Length:     16
           Ratio:   4.83[illegible]              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-817-762-2_COPY_1144_1161 x AF[illegible] ..

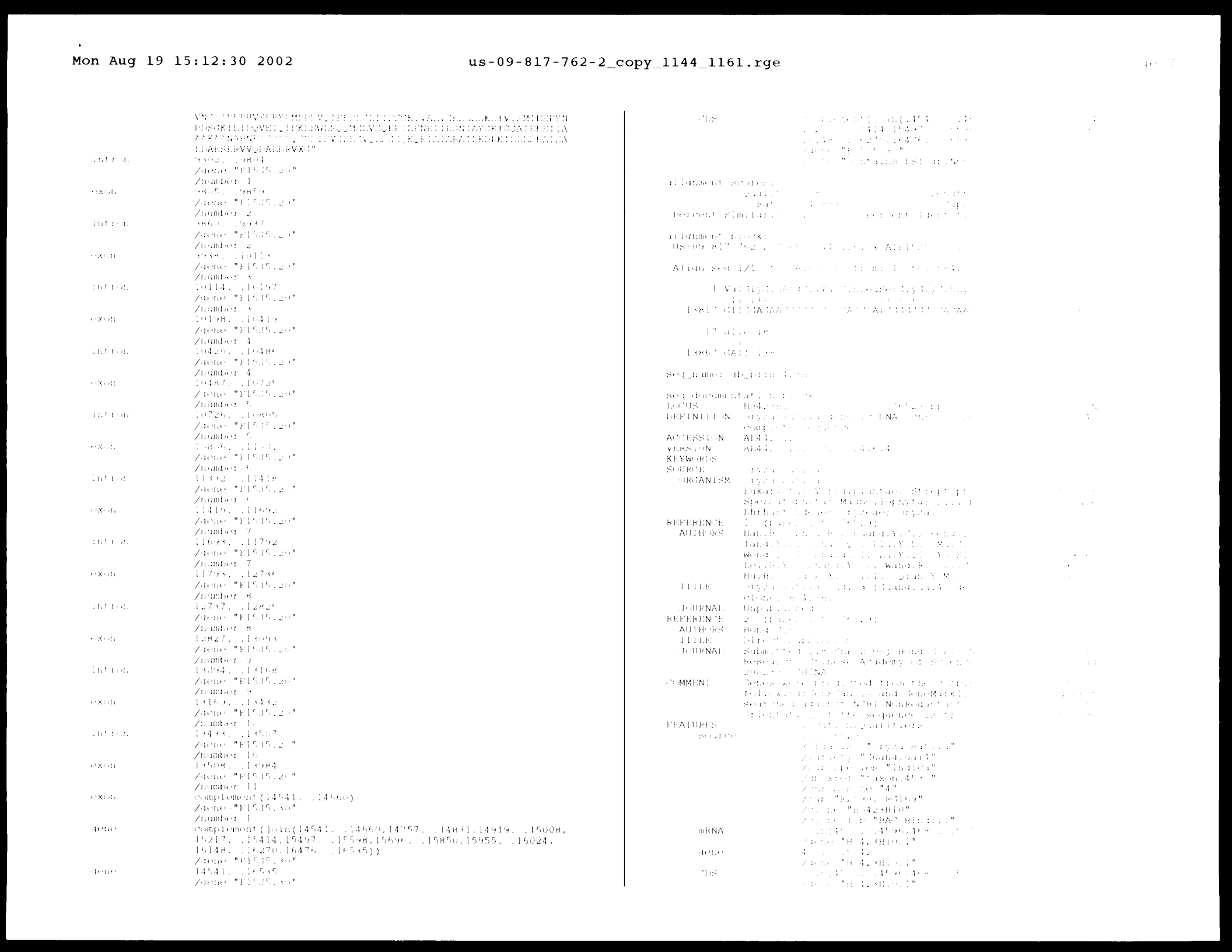
 Align seq 1/1  to reverse of: AB019223  from: 1  to: 51529

      1 ValGlyLeuArgGlyValGluLeuLeuGlyGlyGluLysGluArgIleAl 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
  49120 GT[illegible] [illegible]

     17 aIle 18
        |||
  49070 GAT 49067

seq_name: gb_pl:AF[illegible]

seq_documentation_block:
LOCUS       AF[illegible]                 58427 bp    DNA     linear   PLN 30-AUG-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F15J5 (ESSA
            project).
ACCESSION   AL[illegible]
VERSION     AL[illegible]  GI:5816989
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 58427)
  AUTHORS   Bevan,M., Murphy,G., Ridley,P., Hudson,C., Bancroft,I., Mewes,H.W.,
            Mayer,K.F.X., Lemcke,K. and [illegible]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58427)
  AUTHORS   EU Arabidopsis sequencing project.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-1999) MIPS, at the Max Planck Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK.
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at [illegible]
FEATURES             Location/Qualifiers
     source          1..58427
                     /organism "Arabidopsis thaliana"
                     /variety "Columbia"
                     /db_xref "taxon:3702"
                     /chromosome "4"
     misc_feature    1..2548
                     /note "overlap to BAC [illegible], please refer to [illegible]
                     for analysis and annotation"
     gene            3596..5023
                     /gene "[illegible]"
     exon            complement(3596..3646)
                     /gene "[illegible]"
                     /number 1
     CDS             complement(join(3596..3646,3736..3801,3895..4020,
                     4116..4281,4725..5023))
                     /gene "[illegible]"
                     /note "[illegible] initiation factor 1E signature
                     [DFYCFKEI[illegible]KW]
                     contains EST gb:[illegible], AA[illegible]"
                     /codon_start 1
                     /product "translation initiation factor [illegible]"
                     /protein_id "CAB[illegible]645.1"
                     /db_xref "GI:5816990"
                     /translation "MAVEDIFKGVV[illegible]
                     DGNVD[illegible]
                     NNMK[illegible]
                     [illegible]
                     HEDAKKLDENAKNAYTA"
     gene            complement(join(3596..3646,3736..3801,3895..4020,
                     4116..4281,4725..5023))
                     /gene "F15J5.10"
     intron          complement(3647..3735)
                     /gene "F15J5.10"
                     /number 1
     exon            complement(3736..3801)
                     /gene "F15J5.10"
                     /number 2
     intron          complement(3802..3894)
                     /gene "F15J5.10"
                     /number 2
     exon            complement(3895..4020)
                     /gene "F15J5.10"
                     /number 3
     intron          complement(4021..4115)
                     /gene "F15J5.10"
                     /number 3
     exon            complement(4116..4281)
                     /gene "F15J5.10"
                     /number 4
     intron          complement(4282..4724)
                     /gene "F15J5.10"
                     /number 4
     exon            complement(4725..5023)
                     /gene "F15J5.10"
                     /number 5
     repeat_region   6767..6789
     gene            9077..13984
                     /gene "F15J5.20"
     exon            9077..9301
                     /gene "F15J5.20"
                     /number 1
     CDS             join(9077..9301,9805..9859,9938..10113,10198..10419,
                     10487..10725,10806..11131,11419..11692,11793..12736,
                     12827..13093,13165..13402,13508..13984)
                     /gene "F15J5.20"
                     /note "similarity to multidrug resistance protein 1,
                     Aspergillus fumigatus, U62933
                     Contains ABC transporters family signature,
                     ABC transporter [[illegible]QRIAIARAI][[illegible]QRIAIARAI],
                     ATP/GTP binding site motif A (P-loop), Atp_Gtp_A
                     [GQSGSGKS][GESGSGKS]"
                     /codon_start 1
                     /product "multidrug resistance
                     protein/P-glycoprotein like"
                     /protein_id "CAB[illegible]646.1"
                     /db_xref "GI:5816991"
                     /translation "[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     QIVGA[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     LKNPKI[illegible]IVVVAHRL[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     KVM[illegible]EVSGAGLIQGKAIEG
                     [illegible]AASIFDILDSIPKIDSSS[illegible]
```



```
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09455.1"
                     /db_xref "GI:10241615"
                     /translation "MPKKVHHIESSNS[illegible]
                     [illegible]
                     [illegible]"
     exon            4561..4566
                     /gene "H0423H10.1"
                     /number 1
     intron          4567..4679
                     /gene "H0423H10.1"
                     /number 1
     exon            4680..4742
                     /gene "H0423H10.1"
                     /number 2
     exon            complement(5769..6258)
                     /gene "H0423H10.2"
                     /number 4
     mRNA            complement(join(5769..6258,6542..6546,7862..8061,
                     8089..8148))
                     /gene "H0423H10.2"
     gene            complement(5769..8148)
                     /gene "H0423H10.2"
     CDS             complement(join(5769..6258,6542..6546,7862..8061,
                     8089..8148))
                     /gene "H0423H10.2"
                     /codon_start 1
                     /product "putative cullin 3 protein"
                     /protein_id "CAC09456.1"
                     /db_xref "GI:10241616"
                     /translation "MVA[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]"
     intron          complement(6259..6541)
                     /gene "H0423H10.2"
                     /number 3
     exon            complement(6542..6546)
                     /gene "H0423H10.2"
                     /number 3
     intron          complement(6547..7861)
                     /gene "H0423H10.2"
                     /number 2
     exon            complement(7862..8061)
                     /gene "H0423H10.2"
                     /number 2
     intron          complement(8062..8088)
                     /gene "H0423H10.2"
                     /number 1
     exon            complement(8089..8148)
                     /gene "H0423H10.2"
                     /number 1
     exon            8826..8879
                     /gene "H0423H10.3"
                     /number 1
     mRNA            join(8826..8879,[illegible]..9225,12170..[illegible],12432..12657,
                     12784..12931,13016..13070)
                     /gene "H0423H10.3"
     gene            8826..13070
                     /gene "H0423H10.3"
     CDS             join(8826..8879,[illegible]..9225,12170..[illegible],12432..12657,
                     12784..12931,13016..13070)
                     /gene "H0423H10.3"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09457.1"
                     /db_xref "GI:10241617"
                     /translation "[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     ENKEIP"
     intron          8880..[illegible]
                     /gene "H0423H10.3"
                     /number 1
     exon            [illegible]..9225
                     /gene "H0423H10.3"
                     /number 2
     intron          9226..12169
                     /gene "H0423H10.3"
                     /number 2
     exon            12170..[illegible]
                     /gene "H0423H10.3"
                     /number 3
     intron          [illegible]..12431
                     /gene "H0423H10.3"
                     /number 3
     exon            12432..12657
                     /gene "H0423H10.3"
                     /number 4
     intron          12658..12783
                     /gene "H0423H10.3"
                     /number 4
     exon            12784..12931
                     /gene "H0423H10.3"
                     /number 5
     intron          12932..13015
                     /gene "H0423H10.3"
                     /number 5
     exon            13016..13070
                     /gene "H0423H10.3"
                     /number 6
     exon            complement(15342..15346)
                     /gene "H0423H10.4"
                     /number 2
     mRNA            complement(join(15342..15346,16973..17603))
                     /gene "H0423H10.4"
     gene            complement(15342..17603)
                     /gene "H0423H10.4"
     CDS             complement(join(15342..15346,16973..17603))
                     /gene "H0423H10.4"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09458.1"
                     /db_xref "GI:10241618"
                     /translation "MPNWK[illegible]
                     [illegible]
                     [illegible]
                     [illegible]"
     intron          complement(15347..16972)
                     /gene "H0423H10.4"
                     /number 1
     exon            complement(16973..17603)
                     /gene "H0423H10.4"
                     /number 1
     mRNA            join(20660..20812,20895..[illegible],[illegible]..22324)
                     /gene "H0423H10.5"
     gene            20660..22324
                     /gene "H0423H10.5"
     CDS             join(20660..20812,20895..[illegible],22319..22324)
                     /gene "H0423H10.5"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09459.1"
                     /db_xref "GI:10241619"
                     /translation "MA[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]ARIVKIQDKS"
     exon            20660..20812
                     /gene "H0423H10.5"
```

```
                     /number= 1
     exon            20815..20894
                     /gene="B0423B15.5"
                     /number= 1
     exon            20995..21150
                     /gene="B0423B15.5"
                     /number= 2
     intron          21501..22218

Alignment_Scores:
            Quality:    87.00               Length:      18
              Ratio:  1.833                 Gaps:         0
 Percent Similarity: 100.00%   Percent Identity: 100.00%

Alignment_Block:
 US-09-817-762-2_COPY_1144_1161 x B0423B15 .

 Align Set 1/1   to: B0423B15  from: 1  to: 76529

       1 ValGlyGlnArgGlyValSerIleSerGlyGlyGlnLysGluArgIleAl 17
         |||||| |||| |||||||||||||||||||||||||||||||||||||| 17942
   17942 ValGlyGlnArgGlyValSerIleSerGlyGlyGlnLysGluArgIleAl 17942

      17 aIle 18
         |||
   17943 aIle 17944

Seq Number: US_pl:AP003086

Seq_Documentation_Block:
LOCUS       AP003086                 82430 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MLD15.
ACCESSION   AP003086 BA000014
VERSION     AP003086.1  GI:5672509
KEYWORDS
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MLD15.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. II.
            Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
            and BAC clones
  JOURNAL   DNA Res. 7 (5), 217-221 (2000)
  MEDLINE   20546099
REFERENCE   2  (bases 1 to 82430)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:yuka@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3944)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/arabi/chr3_seq/MLD15
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Information Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/tools.html), GENSCAN (Chris Burge, MIT,
            http://genes.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SpliceProductor (Volker Brendel, Stanford University,
            http://www.fruitfly.org/seq_tools/splice.html).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://rna.wustl.edu/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
```

```
                     /number= 1
     CDS             20815..20894
                     /gene="B0423B15.5"
                     /number= 1
     exon            20995..21150
                     /gene="B0423B15.5"
                     /number= 2
     intron          21501..22218

Alignment_Scores:
            Quality:    87.00               Length:      18
              Ratio:  1.833                 Gaps:         0
 Percent Similarity: 100.00%   Percent Identity: 100.00%

Alignment_Block:
 US-09-817-762-2_COPY_1144_1161 x B0423B15 .

 Align Set 1/1   to: B0423B15  from: 1  to: 76529

       1 ValGlyGlnArgGlyValSerIleSerGlyGlyGlnLysGluArgIleAl 17
         |||||| |||| |||||||||||||||||||||||||||||||||||||| 17942
   17942 ValGlyGlnArgGlyValSerIleSerGlyGlyGlnLysGluArgIleAl 17942

      17 aIle 18
         |||
   17943 aIle 17944

Seq Number: US_pl:AP003086

Seq_Documentation_Block:
LOCUS       AP003086                 82430 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MLD15.
ACCESSION   AP003086 BA000014
VERSION     AP003086.1  GI:5672509
KEYWORDS
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MLD15.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. II.
            Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
            and BAC clones
  JOURNAL   DNA Res. 7 (5), 217-221 (2000)
  MEDLINE   20546099
REFERENCE   2  (bases 1 to 82430)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:yuka@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3944)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/arabi/chr3_seq/MLD15
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Information Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/tools.html), GENSCAN (Chris Burge, MIT,
            http://genes.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SpliceProductor (Volker Brendel, Stanford University,
            http://www.fruitfly.org/seq_tools/splice.html).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://rna.wustl.edu/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
```

```
                     [illegible] 42064..42139, ..42181)
                     /note="gene_id:MLD[illegible]
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB0[illegible].1"
                     /db_xref="GI:9294[illegible]"
                     /translation="[illegible]
                     KS[illegible]SYAKR[illegible]
                     [illegible]
                     [illegible]
                     EAV[illegible]VQAG
                     DEAVLELEINENPDVNDKN[illegible]
                     [illegible]
                     VE[illegible]
                     [illegible]ANWK[illegible]MKV
                     [illegible]
                     [illegible]ASYNNN[illegible]
                     [illegible]
                     [illegible]NNKMIYRGI"
     CDS             complement(42992..45127)
                     /note="gene_id:MLD15.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="disease resistance protein-like"
                     /protein_id="BAB02[illegible].1"
                     /db_xref="GI:9294[illegible]"
                     /translation="MK[illegible]
                     LLKPK[illegible]
                     S[illegible]SYLHGKEHSN[illegible]
                     V[illegible]
                     [illegible]NLEVNN[illegible]
                     [illegible]
                     LRSLE[illegible]LSDNNF[illegible]
                     N[illegible]
                     [illegible]
                     K[illegible]VESEI[illegible]
                     LIALESLDVSQNKL[illegible]
                     N[illegible]EEDEEVISWIAAA[illegible]
                     Y[illegible]VEYKP[illegible]WEIKIF[illegible]NN[illegible]RRS[illegible]H"
     CDS             complement(join(49[illegible]..49243,49323..49452,50436..[illegible]
                     50623..50701))
                     /note="gene_id:MLD15.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="60S ribosomal protein L34"
                     /protein_id="BAB0[illegible].1"
                     /db_xref="GI:92942[illegible]"
                     /translation="MV[illegible]
                     PV[illegible]
                     VKKVLKI[illegible]KAKEKIAPK[illegible]"
     CDS             join(5[illegible]29..57[illegible]61,58188..58[illegible]17,58434..5[illegible]42)
                     /note="gene_id:MLD15.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="MYB family transcription factor-like protein"
                     /protein_id="BAB02134.1"
                     /db_xref="GI:92942[illegible]"
                     /translation="MVR[illegible]K[illegible]VKK[illegible]PWIPEE[illegible]NWKAVP[illegible]
                     [illegible]SKS[illegible]RLRW[illegible]
                     [illegible]KNYWN[illegible]KKKLNKVN[illegible]
                     [illegible]AKKAL[illegible]
                     [illegible]
                     SP[illegible]
                     [illegible]
     CDS             complement(63869..69193)
                     /note="gene_id:MLD15.[illegible]"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
                     /product="non-LTR retroelement reverse transcriptase-like
                     protein"
     CDS             71[illegible]..71602
                     /note="gp:AAF[illegible]
```

```
                     gene_id:MLD15.1[illegible]
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB02135.1"
                     /db_xref="GI:9294233"
                     /translation="[illegible]
                     [illegible]N[illegible]N"
     CDS             complement(74[illegible]..7[illegible]14)
                     /note="contains similarity to retrotransposon reverse
                     transcriptase"

alignment scores:
          Quality:    67.00                Length:      18
            Ratio:    4.[illegible]                  Gaps:
 Percent Similarity: [illegible]    Percent Identity: [illegible]

alignment block:
 US-09-817-762-2_COPY_1144_1161 x AP060386/rev [illegible]

 Align seq 1/1 to: reverse of: AP060386 from: 1 to: 82430

        [illegible] Val[illegible]
        [illegible]
   2[illegible] AA[illegible]

       17 aile 18
          |-|:
   19965 AAT 19962

seq name: gb_pl:AB026644

 seq_documentation_block:
LOCUS       AB026644                85430 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MLD20.
ACCESSION   AB026644 BA000014
VERSION     AB026644.1  GI:4757400
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MLD20
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. I.
            Sequence features of the regions of 4,504,864 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (2), 131-135 (2000)
  MEDLINE   20277480
REFERENCE   2  (bases 1 to 85430)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://[illegible]
            Genes with similarity to proteins in the databases are described as
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://[illegible]) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://[illegible]).
            Genes encoding tRNAs are predicted by tRNAscan-SE
```

```
                (Genbank, Swissprot & Geneseq data is extracted from CDS
                translations and also from GenPept, Swissprot & PIR. In
                addition Sequences may not be the public ones of this status. It may be
                similar because we test for overlaps between neighboring submissions.
                The "*" status is MAYBE and the "!" status is ZERO.
FEATURES        Location/Qualifiers
     source     1..1854
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /chromosome="3"
                /note="MBK5_5"
                /clone_lib="M. Sussman"
     misc_feature
                complement(1..1854)
                /note="CDS is reported to Acc# AAC49481"
                gene_id=MZB5.5"
                /number=1
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(1854..2804)
                /note="CDS is reported to Acc# AAC49482"
                gene_id=MZB5.6"
                /number=2
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(2804..3898)
                /note="CDS is reported to Acc# AAC49483"
                gene_id=MZB5.7"
                /number=3
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(3898..4224)
                /note="CDS is reported to Acc# AAC49484"
                gene_id=MZB5.8"
                /number=4
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                join(14479..14561,14741..14900,15053,15275..15345,
                15545..15737,15762..15824)
                /note="similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /product="unknown"
                /db_xref="GI:2494"
                /translation="MAVIIKKIKDFISERSFILKVVGLKKLIGLGSLVEA
                ANGTVLHAKEIKYAIEGLNTTEVLSLNGSIISKFLVIIINGVVALKKVYKEN
                SKIS_IKKVIEYSLKFFLMMLIRLSVLVKKEFKIVFSGLAIEKTGIIANIKAIL
                LGSGLIEAIKNLISLVHSKLAKFLLYLAFSEMLRLIIHIMKHENGLYRILYM"
     misc_feature
                complement(join(15479..15561,15741..15900,16053,16275..
```

```
                (Genbank, Swissprot & Geneseq data is extracted from CDS
                translations and also from GenPept, Swissprot & PIR. In
                addition Sequences may not be the public ones of this status. It may be
                similar because we test for overlaps between neighboring submissions.
                The "*" status is MAYBE and the "!" status is ZERO.
FEATURES        Location/Qualifiers
     source     1..1854
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /chromosome="3"
                /note="MBK5_5"
                /clone_lib="M. Sussman"
     misc_feature
                complement(1..1854)
                /note="CDS is reported to Acc# AAC49481"
                gene_id=MZB5.5"
                /number=1
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(1854..2804)
                /note="CDS is reported to Acc# AAC49482"
                gene_id=MZB5.6"
                /number=2
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(2804..3898)
                /note="CDS is reported to Acc# AAC49483"
                gene_id=MZB5.7"
                /number=3
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                complement(3898..4224)
                /note="CDS is reported to Acc# AAC49484"
                gene_id=MZB5.8"
                /number=4
                /evidence=not_experimental
                /product="multidrug resistance polyprotein"
     misc_feature
                join(14479..14561,14741..14900,15053,15275..15345,
                15545..15737,15762..15824)
                /note="similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /product="unknown"
                /db_xref="GI:2494"
                /translation="MAVIIKKIKDFISERSFILKVVGLKKLIGLGSLVEA
                ANGTVLHAKEIKYAIEGLNTTEVLSLNGSIISKFLVIIINGVVALKKVYKEN
                SKIS_IKKVIEYSLKFFLMMLIRLSVLVKKEFKIVFSGLAIEKTGIIANIKAIL
                LGSGLIEAIKNLISLVHSKLAKFLLYLAFSEMLRLIIHIMKHENGLYRILYM"
     misc_feature
                complement(join(15479..15561,15741..15900,16053,16275..
```

```
                     VIGKGKIVESGIHSSLLEKGPIGIYFSLAGIQRIL*"
     CDS             complement(join(33545..34937,35632..35874,36668..36788,
                     36787..37042,37125..37305,37436..37499,37887..38087))
                     /note="gene F13M22.17"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="multidrug resistance p-glycoprotein, ABC
                     transporter like protein"
                     /protein_id="BAB02[illegible].1"
                     /db_xref="GI:[illegible]"
                     /translation="MMKS[illegible]
                     SKLNNV[illegible]FFM[illegible]VARNAVALYVVA[illegible]AAKK
                     GKYIKAVIT[illegible]HVIT[illegible]SVIIVVG[illegible]SKLPNFLMNIGAIVASY
                     IV[illegible]WFL[illegible]IV[illegible]FGLMYIRALISISMK[illegible]YNEAGSIADQVLSSVR

Alignment Scores:
            Quality:    8[illegible]             Length:      18
              Ratio:    4.[illegible]             Gaps:         [illegible]
 Percent Similarity: 100.000    Percent Identity: 100.000

Alignment Block:
 US-09-817-762-2_COPY_1144_1161 x AB026644/rev ..

 Align seq 1/1 to: reverse of: AB026644 from: 1 to: 85436

        [illegible] ValGly[illegible]ArgGlyValIleLeuSerGlyLysGlnLys[illegible]ArgIleA[illegible]
          [illegible]
    85[illegible] [illegible]AAA[illegible]A[illegible]AA[illegible]AAAA[illegible]AGGATI[illegible] [illegible]

        [illegible]

    [illegible] AA[illegible]

Sequence: [illegible] AC[illegible]

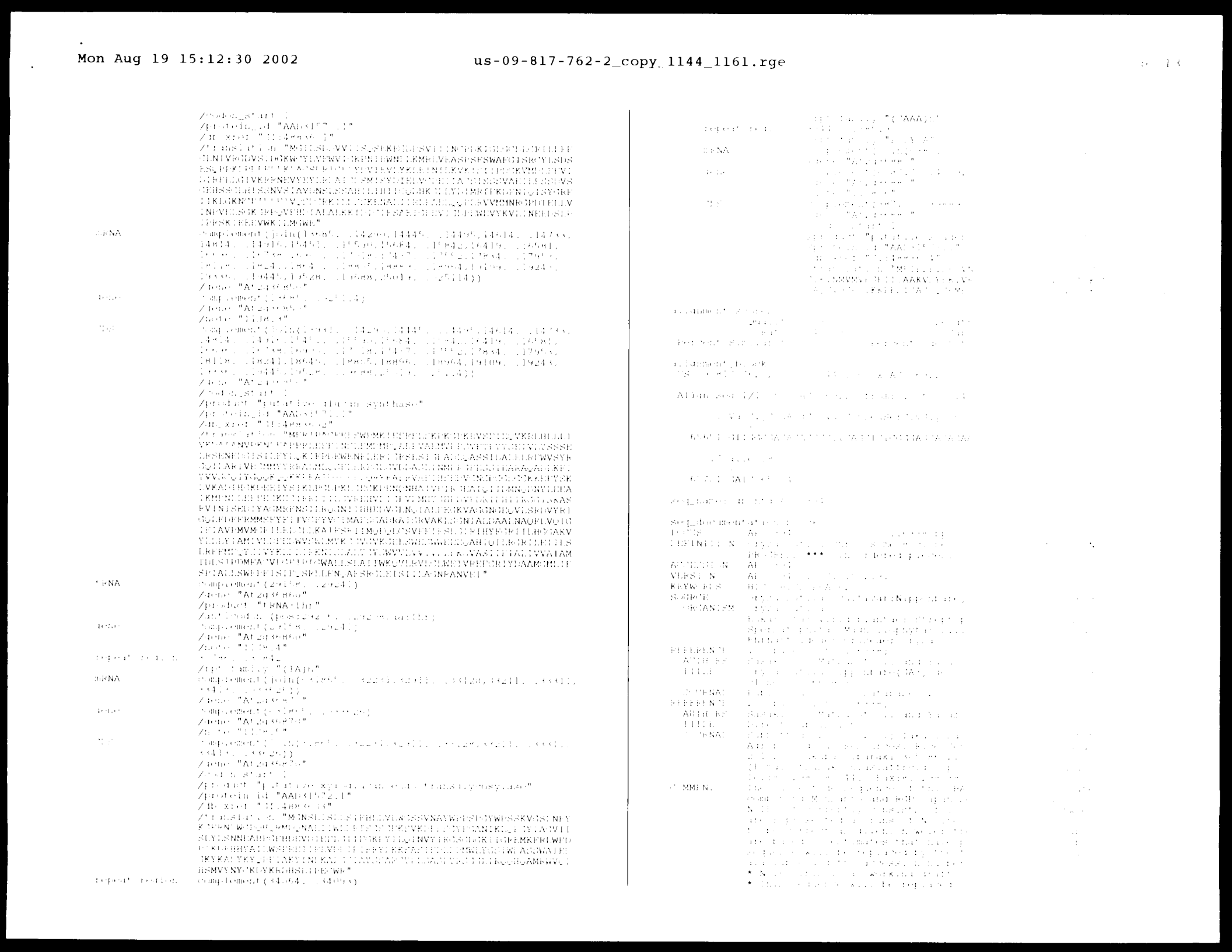
seq_documentation_block:
LOCUS       AC006922              114960 bp    DNA     linear   PLN 05 APR 2000
DEFINITION  Arabidopsis thaliana chromosome II section 201 of 255 of the
            complete sequence. Sequence from clones T1J8.
ACCESSION   AC006922 AE002093
VERSION     AC006922.6  GI:6598638
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 114960)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083[illegible]
   PUBMED   10617197
REFERENCE   2  (bases 1 to 114960)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:[illegible].
            The sequence and annotation of this region were merged from those
            of the individual clones on this chromosome after resolving
            overlaps. For detailed information please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (http://compbio.ornl.gov/tools/grail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://genes.mit.edu/GENSCAN.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetPGene/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on these analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Those without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan SE (Sean Eddy,
            http://[illegible]). Repeats were
            identified by repeatmasker (Arian Smit,
            http://[illegible]/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
            F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
            F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Bruce Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to: [illegible]
FEATURES             Location/Qualifiers
     source          1..114960
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     misc_feature    1..114960
                     /note="Sequence from clone T1J8"
     mRNA            join(4393..4869,4961..5526)
                     /gene="At2g36830"
     gene            4393..5526
                     /gene="At2g36830"
                     /note="T1J8.1"
     misc_feature    4437..4813
                     /gene="At2g36830"
                     /note="molecular marker [illegible]"
     CDS             join(448[illegible]..48[illegible],4961..53[illegible])
                     /gene="At2g36830"
                     /codon_start=1
                     /product="putative aquaporin (tonoplast intrinsic protein
                     gamma)"
                     /protein_id="AAD3[illegible].1"
                     /db_xref="GI:4883[illegible]"
                     /translation="M[illegible]AAL[illegible]
                     NKLTENGA[illegible]VAAAVAHAF[illegible]VAV[illegible]GANISG[illegible]VNLAVI[illegible]GAF[illegible]GNI[illegible]
                     [illegible]
                     [illegible]
                     YWAGPLVGG[illegible]IAGL[illegible]YEV[illegible]INT[illegible]HEQ[illegible]DY"
     repeat_region   5766..5809
                     /rpt_family="PolyA"
     mRNA            complement(join(10830..11067,11066..11[illegible],11[illegible]..11[illegible],
                     11871..11970,12054..12281,12[illegible]..12546))
                     /gene="At2g36840"
     gene            complement(10830..12546)
                     /gene="At2g36840"
                     /note="T1J8.2; predicted by genefinder"
     CDS             complement(join(1[illegible]..1[illegible],11[illegible]..11[illegible],11[illegible]..11[illegible],
                     11871..11970,12054..12281,12[illegible]..12546))
                     /gene="At2g36840"
                     /note="hypothetical protein"
```



```
            * By the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126786
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OJ1066_B05"
gap                  complement(<1..>2)
                     /estimated_length=unknown
                     /gap_type="unknown"
                     /linkage_evidence="unspecified"
ORIGIN


//

LOCUS       AL606656             126786 bp    DNA     linear   HTG 28-SEP-2002
DEFINITION  Oryza sativa chromosome 8 clone OSJNBb0079B02, *** SEQUENCING IN
            PROGRESS ***, 16 ordered pieces.
ACCESSION   AL606656
VERSION     AL606656.1  GI:32394009
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            BEP clade; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 126786)
  AUTHORS   Rao,G., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Ma,L., Yu,Z., Chen,L., Fan,D.L., Wang,G.C., Zhang,L.,
            Lu,Y.Q., Su,S.C., Liu,C.Y., Lu,Z.B., Du,J.J., Zhang,Y.J., Gu,Y., Li,C.,
            Zhang,Y., Hu,H., Jiang,X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Han,B., Zhang,L., Wu,M., Zhang,K.Q., Guan,J.P., Fu,G.,
            Wang,S.T., Xu,X.X., Liu,F., Liu,W., Gu,W.Y., Zhu,G.F., Li,Y.F.,
            Jiang,L., Yin,H.F., Zhang,Y., Sai,Z., Chen,C., Kang,H., Chen,X.Y.,
            Shao,Y.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.D.,
            Ding,C.W., Shen,H.H., Gu,J.L., Chen,S.L., Ni,L., Zhu,F.H. and
            Huang,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2002) Rao Bin, National Center for Gene Research,
            Chinese Academy of Sciences, 300# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (cultivar-group) genomic DNA, chromosome 8, BAC
            clone: OSJNBb0079B02.
COMMENT     Web site: http://www.ncgr.ac.cn
                        Summary Statistics
            Assembly program: phrap
                        .
            This is a complete sequence.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126786
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OJ1066_B05"
gap                  complement(<1..>2)
                     /estimated_length=unknown
                     /gap_type="unknown"
                     /linkage_evidence="unspecified"
ORIGIN


//

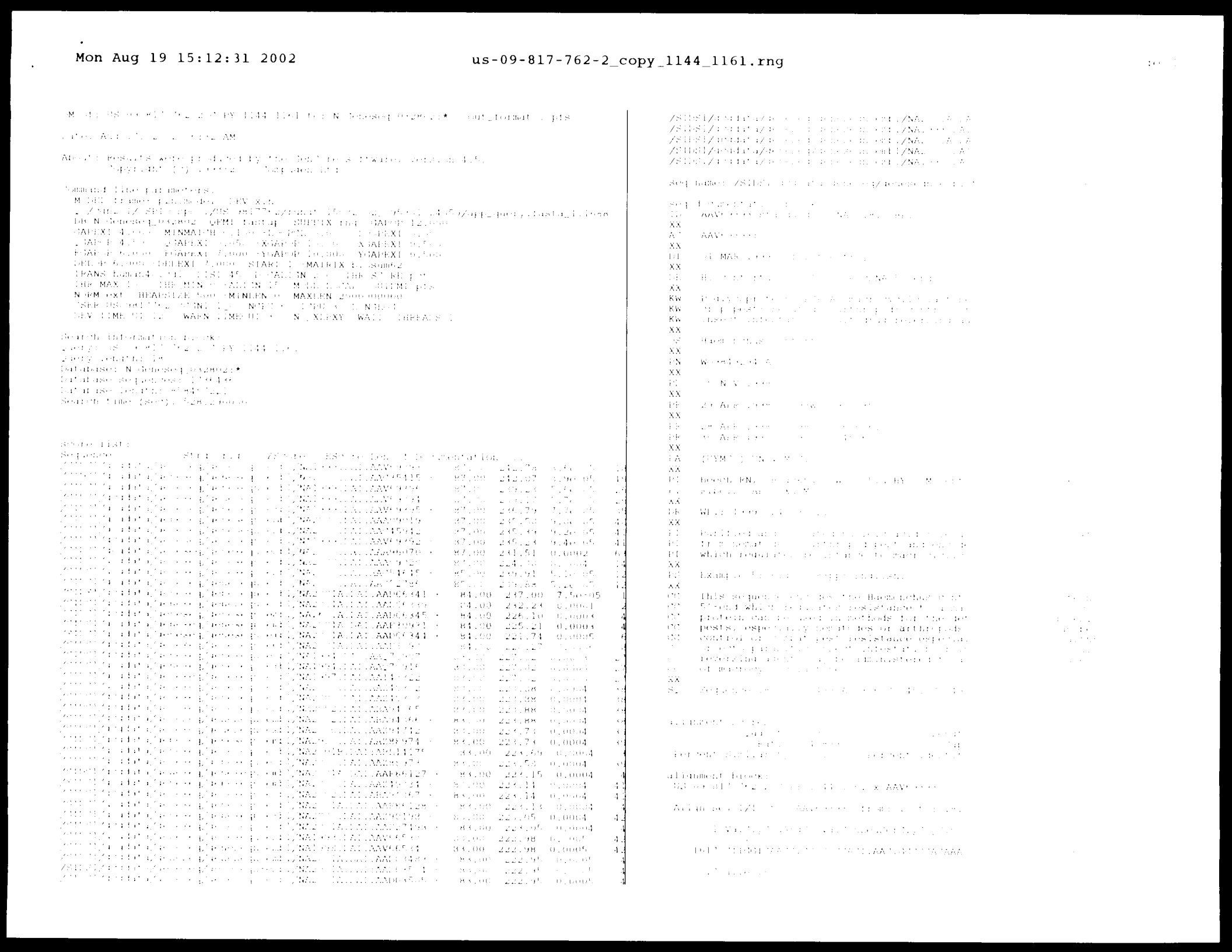
LOCUS       AL606656             126786 bp    DNA     linear   HTG 28-SEP-2002
DEFINITION  Oryza sativa chromosome 8 clone OSJNBb0079B02, *** SEQUENCING IN
            PROGRESS ***, 16 ordered pieces.
ACCESSION   AL606656
VERSION     AL606656.1  GI:32394009
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            BEP clade; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 126786)
  AUTHORS   Rao,G., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Ma,L., Yu,Z., Chen,L., Fan,D.L., Wang,G.C., Zhang,L.,
            Lu,Y.Q., Su,S.C., Liu,C.Y., Lu,Z.B., Du,J.J., Zhang,Y.J., Gu,Y., Li,C.,
            Zhang,Y., Hu,H., Jiang,X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Han,B., Zhang,L., Wu,M., Zhang,K.Q., Guan,J.P., Fu,G.,
            Wang,S.T., Xu,X.X., Liu,F., Liu,W., Gu,W.Y., Zhu,G.F., Li,Y.F.,
            Jiang,L., Yin,H.F., Zhang,Y., Sai,Z., Chen,C., Kang,H., Chen,X.Y.,
            Shao,Y.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.D.,
            Ding,C.W., Shen,H.H., Gu,J.L., Chen,S.L., Ni,L., Zhu,F.H. and
            Huang,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2002) Rao Bin, National Center for Gene Research,
            Chinese Academy of Sciences, 300# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (cultivar-group) genomic DNA, chromosome 8, BAC
            clone: OSJNBb0079B02.
COMMENT     Web site: http://www.ncgr.ac.cn
                        Summary Statistics
            Assembly program: phrap
                        .
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126786
                     /organism="Oryza sativa"
```

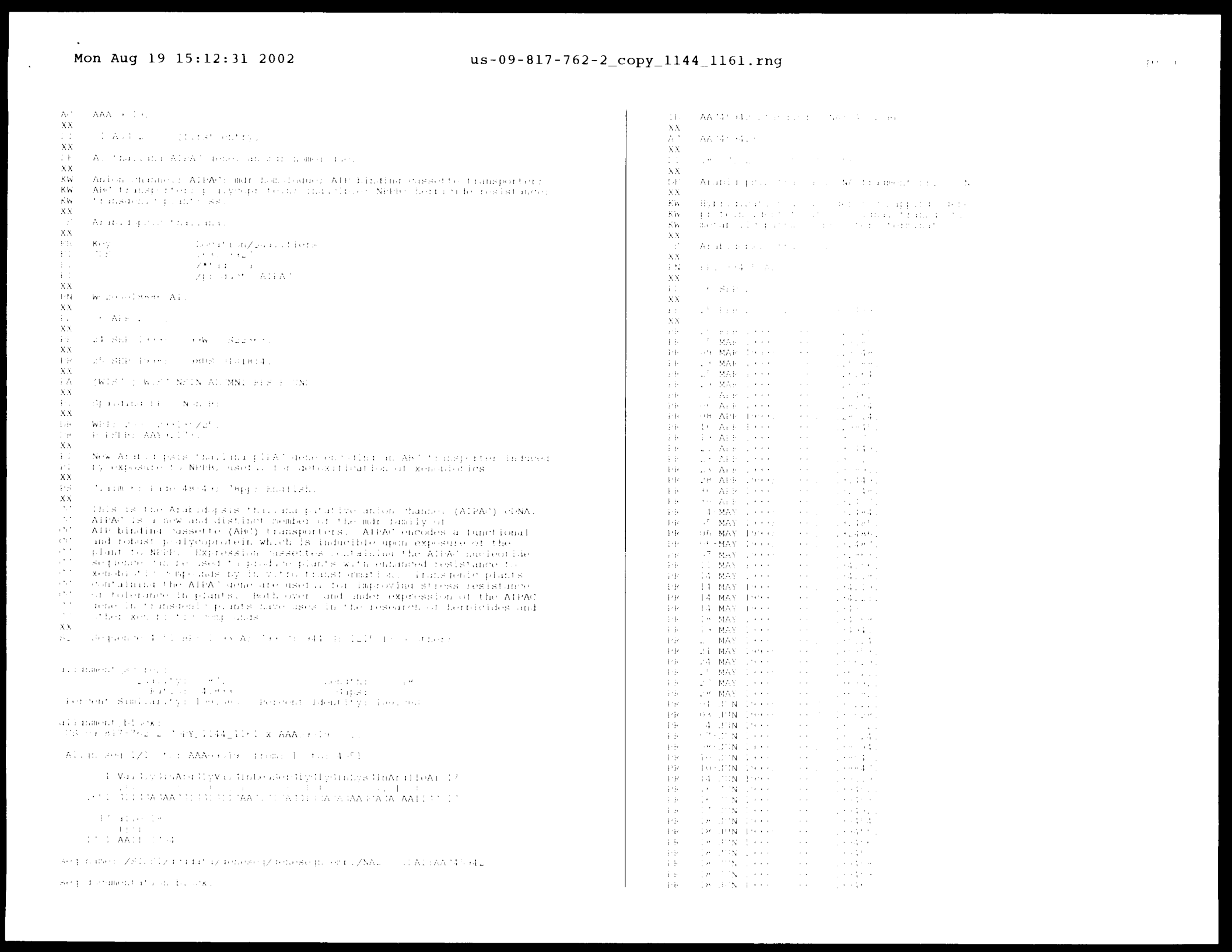
```
                     /number= 5
     exon            1683..1703
                     /gene="AI4q17700"
                     /number=6
     misc_feature    complement(join(1683..1703,1744..1766,1814..1866,
                     1903..1965,2016..2090,2148..2187,2255..2299,2357..
                     2441..2464,2488..2507))
                     /gene="AI4q17700"
     CDS             complement(join(1683..1703,1744..1766,1814..1866,
                     1903..1965,2016..2090,2148..2187,2255..2299,2357..
                     2441..2464,2488..2507))
                     /gene="AI4q17700"
                     /note="similarity to Arabidopsis
                     thaliana
                     sequence [gb|AB007741] GLABRA2 like protein
                     (GLABRA2) gene, complete cds. ESTs
                     gb|AA712144 gb|AA585268 and gb|AA585269 come
                     from this gene"
                     /codon_start=1
                     /product="GLABRA2 like protein"
                     /protein_id="AAB07741.1"
                     /db_xref="GI:2760741"
                     /translation="MEEKGEKKGKSHDVLNGLNVEGGVSSSPCTTLNFNYTTSELNEN
                     SPYTPLPSSSSSMVSPCSLSPSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     MELALPKENAHPDDKTPKLPLSKKLTLSPTLVKEWFLNKPTIKAIGPSFNAKLKARNE
                     TLKHESLNTLSNESLPSPSSPDHNLESLNADDLPLLSLSCTVDMNSSSSPCTTPPPI
                     NKNNNLNLLAELKAIMELAVSPPSAKMTLNLSLGNKPSSSSSCVTLSEESYK
                     RMFGCLNMLTSPTTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     ISSTASSSSSLLSLNAISSKWMVVFPPLLKEASALTLLYKKKPSGTLQAMSNT
                     YSLVTWFPVFEERPLFPVNFFVDLYSLSLMNLGLLSLNITLK
                     LSNIYLLPFLMLPSVSAFKNLMKLSLPVKTTLNSLNCLSLAPTKLTVKVSPKVT
                     TEVLTAVSVLLSYSSLSLVFLLKLQNLSLSVFSNSSSNVELMLSLDLTLNSSSLLVY
                     SLVSLVSLSLSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     SNVLTERCLSLSVSVINHPLTALVNELLSALVNLVSN"
     exon            complement(1667..1741)
                     /gene="AI4q17710"
                     /number=1
     misc_feature    complement(1667..1741)
                     /gene="AI4q17710"
     CDS             complement(1667..1741)
                     /gene="AI4q17710"
                     /note="similarity to Arabidopsis
                     thaliana
                     sequence [gb|AB007741] GLABRA2 like protein
                     (GLABRA2) gene, complete cds. ESTs
                     gb|AA712144 gb|AA585268 and gb|AA585269 come
                     from this gene"
                     /codon_start=1
                     /product="GLABRA2 like protein"
                     /protein_id="AAB07741.1"
                     /db_xref="GI:2760741"
     misc_feature    complement(1742..1766)
                     /gene="AI4q17710"
                     /number=2
     exon            complement(1742..1766)
                     /gene="AI4q17710"
                     /number=2
     misc_feature    complement(1814..1866)
                     /gene="AI4q17710"
                     /number=3
     misc_feature    complement(1836..1866)
                     /gene="AI4q17710"
                     /number=4
     misc_feature    complement(1903..1965)
                     /gene="AI4q17710"
                     /number=5
     misc_feature    complement(1904..1965)
                     /gene="AI4q17710"
                     /number=6
     misc_feature    complement(1983..2007)
                     /gene="AI4q17710"
                     /number=7
     exon            complement(1983..2090)
                     /gene="AI4q17710"
                     /number=8
     misc_feature    complement(2016..2090)
                     /gene="AI4q17710"
                     /number=9
     misc_feature    complement(2148..2187)
                     /gene="AI4q17710"
                     /number=10
     exon            complement(2148..2299)
                     /gene="AI4q17710"
                     /number=11
     misc_feature    complement(2255..2299)
                     /gene="AI4q17710"
                     /number=12
     misc_feature    complement(2357..2441)
                     /gene="AI4q17710"
                     /number=13
     misc_feature    complement(2357..2464)
                     /gene="AI4q17710"
                     /number=14
     misc_feature    complement(2488..2507)
                     /gene="AI4q17710"
                     /number=15
     misc_feature    complement(2488..2507)
                     /gene="AI4q17710"
                     /number=16
     misc_feature    complement(2488..2507)
                     /gene="AI4q17710"
                     /number=17
     exon            complement(2488..2507)
                     /gene="AI4q17710"
                     /number=18
```

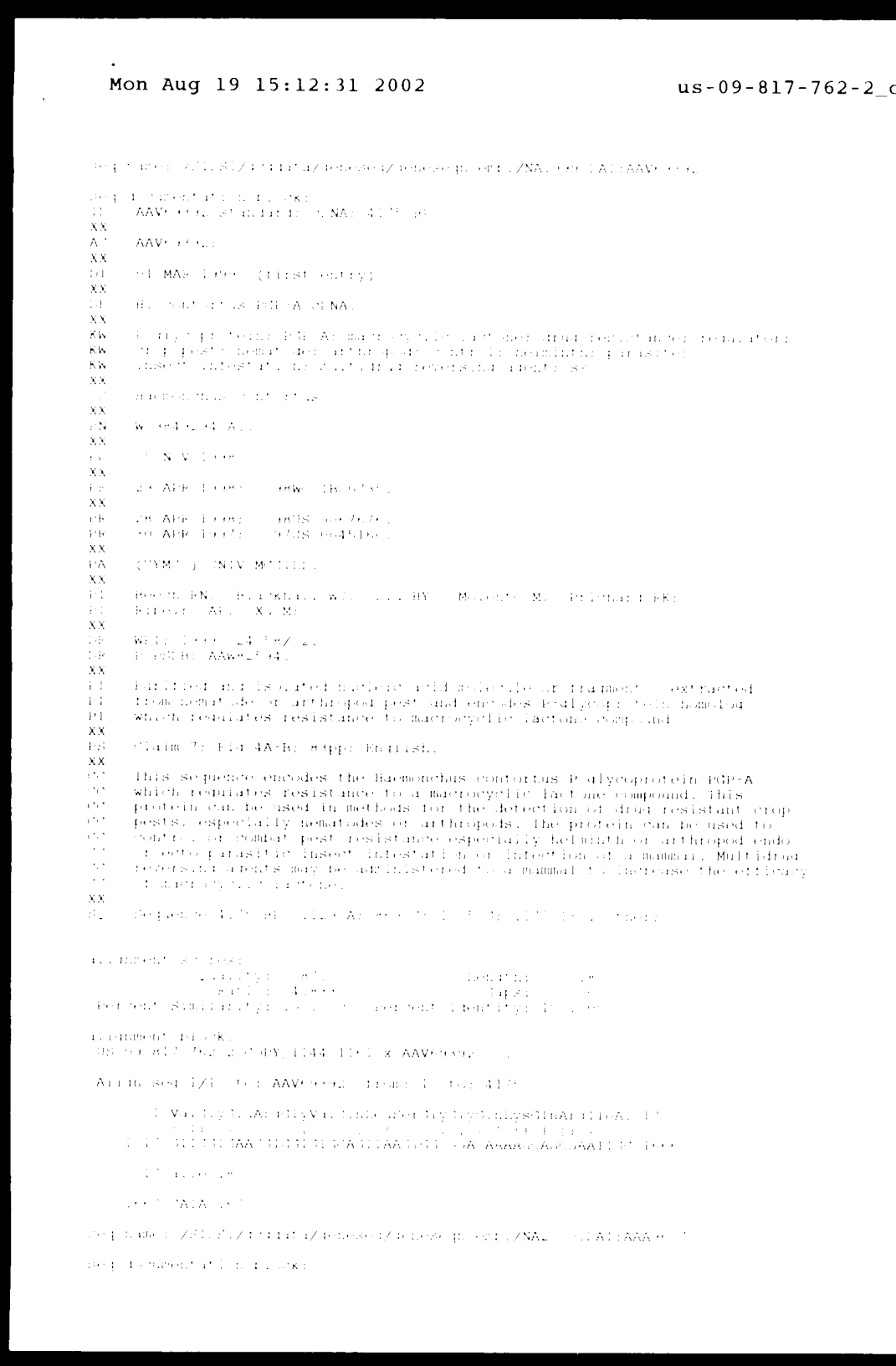
```
     misc_feature    complement(1934..1944)
                     /gene="AI4q17710"
                     /number=19
     exon            complement(1941..1944)
                     /gene="AI4q17710"
                     /number=20
     misc_feature    complement(1945..1967)
                     /gene="AI4q17710"
                     /number=21
     exon            complement(1945..1967)
                     /gene="AI4q17710"
                     /number=22

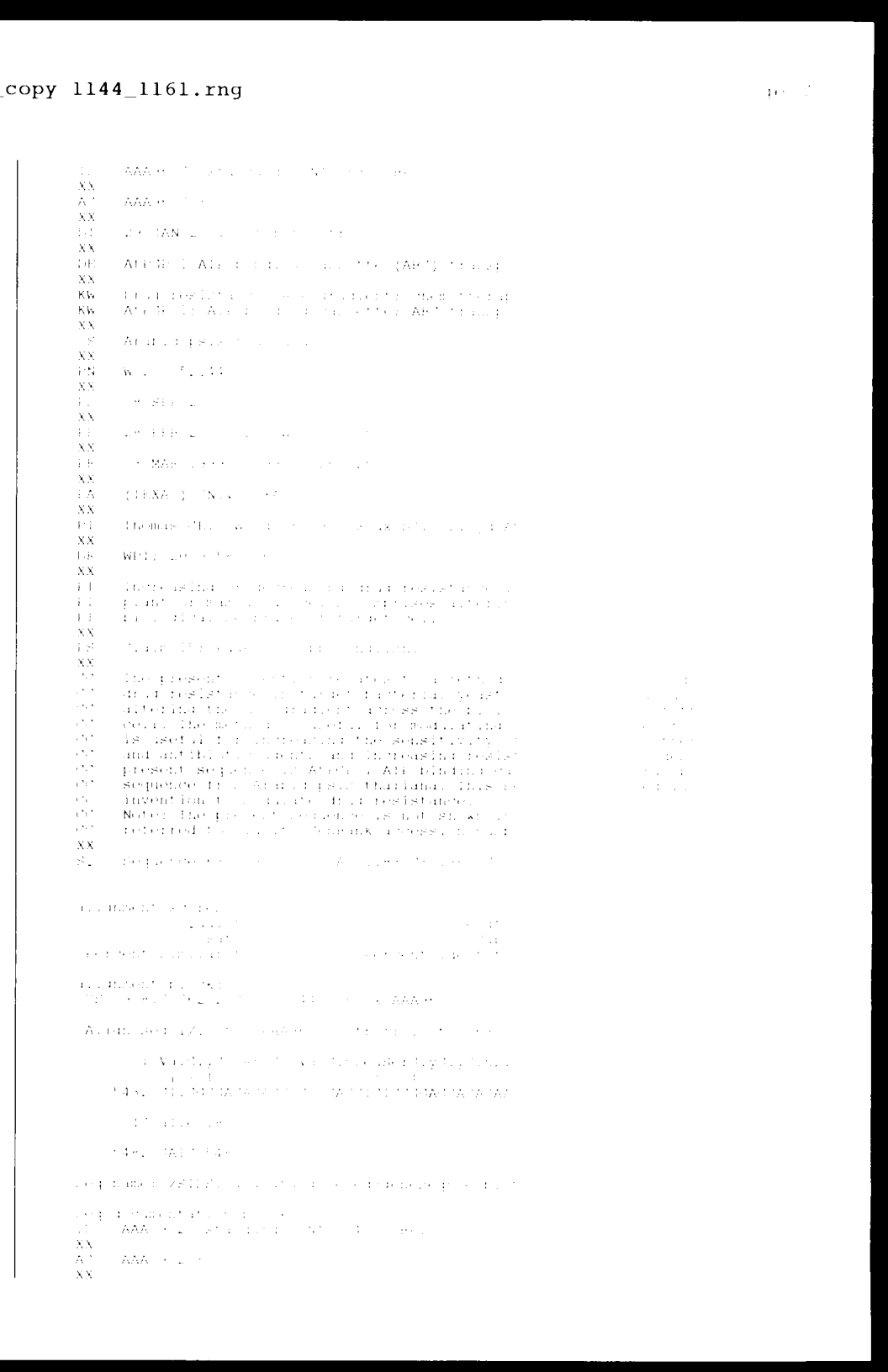
BASE COUNT  [illegible]
ORIGIN
        1 [illegible]
//

LOCUS       AIPGLY2                    9924 bp    mRNA    linear   PLN 15-MAY-2001
DEFINITION  Arabidopsis mRNA for Proglycoprotein 2.
ACCESSION   Y[illegible]
VERSION     Y[illegible]  GI:[illegible]
KEYWORDS    [illegible] pgp2 gene.
SOURCE      Arabidopsis thaliana
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophytes;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 9924)
  AUTHORS   Marques,J.P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9924)
  AUTHORS   Marques,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-1996) J.P. Marques, Max Planck Institut fuer
            Zuechtungsforsch, Genetic Principles of Plant Breeding,
            Carl von Linne Weg 10, 50829 Koeln, DE
FEATURES             Location/Qualifiers
     source          1..9924
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /map="4 cM"
                     /clone_lib="[illegible]"
     gene            [illegible]
                     /gene="pgp2"
     CDS             [illegible]
                     /gene="pgp2"
                     /codon_start=1
                     /product="P-glycoprotein 2"
                     /protein_id="CAA71277.1"
                     /db_xref="GI:2760842"
                     /db_xref="SPTREMBL:O04211"
                     /translation="MQDSHPAPCKEEMLPKVSLLKLFSFADTVDVMLMVVGT
                     LGDVASGVSIPLFFFGKLINSLGYAYLFPDKSHRVAKYSLDFVYLGVAIFSSWLEVA
                     WMHTGERQAAKMRMAYLRSMLSQDVSFFDTDVRTSEVISAISSDSLVVQDAISEKV
                     NKFLHYLSTFVAGFVIGFVRGWQLTLVILAVSPVLGLSAAVWAKILSSFTDKELLA
```









```
;      ( A ' .      (file' entry)
XX
;;    A. thaliana AtAA' nucleic DNA.
XX
RN    AC.  ( Stable ) AtAA' ATP binding cassette transporter
RN    AA' transporter ( e.g. pseudomonas) detoxifies NEMs herbicide resistance
RN    transgenic plants.
XX
 (   Arabidopsis thaliana.
XX
FH    Key               Location/Qualifiers
;     [comment          1..582
;                       /*tag= a
;                       /note= "promoter translation region."
;      .5               <1..>1
;                       /*tag= i
;     CDS               <1..>582
;                       /*tag= c
;                       /number= 1
;     /gene=            "AtAA'"
;                       /*tag= d
;                       /number= 1
;     /note             "ATP binding
;                       /*tag= e
;                       /number= 1
;     /codon_start=     1
;                       /*tag= f
;                       /number= 1
;     /transl_table=    11
;                       /*tag= g
;                       /number= 1
;     /product=         "AtAA' protein"
;                       /*tag= h
;                       /number= 1
;     /protein_id=      "AAA'"
;                       /*tag= k
;                       /number= 1
;     /db_xref=         "GI:"
;                       /*tag= l
;                       /number= 1
;     /translation=     "
;                       /*tag= m
;                       /number= 1
;     BASE COUNT        
;                       /*tag= s
;                       /number= 1
;     ORIGIN            
;                       /*tag= o
;                       /number= 1
FH                      /*tag= b
FT                      /number= 1
FT    CDS               
FT                      /*tag= c
FT                      /number= 2
FT    /gene=            
FT                      /*tag= d
FT                      /number= 2
FT    /note             
FT                      /*tag= e
FT                      /number= 2
FT    /codon_start=     
FT                      /*tag= f
FT                      /number= 2
FT    /transl_table=    
FT                      /*tag= g
FT                      /number= 2
FT    /product=         
FT                      /*tag= h
FT                      /number= 2
FT    /protein_id=      
FT                      /*tag= k
FT                      /number= 2
FT    /db_xref=         
FT                      /*tag= l
FT                      /number= 2
FT    /translation=     
FT                      /*tag= m
FT                      /number= 2
FT    BASE COUNT        
FT                      /*tag= s
FT                      /number= 2
FT    ORIGIN            
FT                      /*tag= o
FT                      /number= 2
XX
EN    Wed Aug 14 2002 A..
XX
RC    * AtA gene *
XX
RI    24-SEP-1998;    cDW: 05x2x96.
XX
RE    25-SEP-1998;    9805-GIDIB14
XX
RA    (WISC ) WISC NCSN ALUMNI RES FOUND
XX
RT    Specific RT:  N.A.B.
XX
RP    WPI: ... /...
RP    P-US-5: AAY/C1''
XX
RT    New Arabidopsis thaliana plAA' gene secretase in AA' transporter related
RT    DNA sequence for AtAA, useful for detoxification of xenobiotics
XX
RL    Claim 6; Page 70-74; 78pp; English.
XX
CC    This is the Arabidopsis thaliana potentiation channel (AtPA') protein
CC    DNA.  AtPA' is a new and distinct member of the ATP family of
CC    ATP binding cassette (ABC) transporters.  AtPA' encodes a functional
CC    and robust plant protein which is inducible upon exposure of the
CC    plant to NEMs.  Expression cassettes containing the AtPA' protein
CC    sequence can be used to produce plants with enhanced resistance to
CC    xenobiotic compounds by in vitro transformation.  Transgenic plants
CC    containing the AtPA' gene are useful for improving stress resistance
CC    or tolerance in plants.  Both over- and under expression of the AtPA'
CC    gene in transgenic plants have uses in the research of herbicides and
CC    other xenobiotic compounds.
XX
SQ    Sequence 11 ... BP; 44% A; 22% ... 22% G; 4% ... other;


Alignment_Number 2 Score:
                 Query:  1 ...                  Length:  ...
                   Match:  ...                    Gaps:
 Percent Similarity: 100.00%   Percent Identity: 100.00%

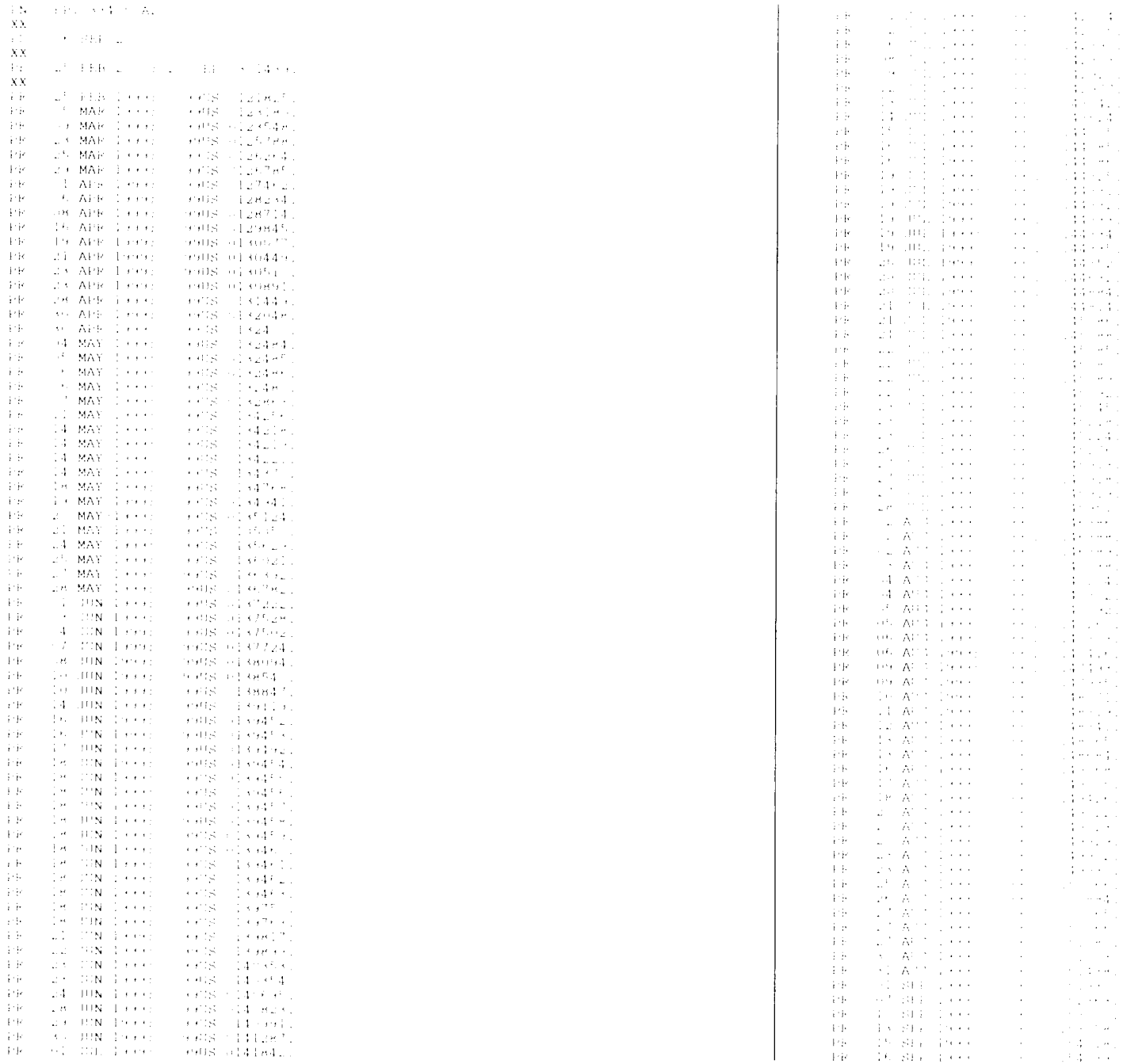
Alignment_blocks:
 US-09-817-762-2_COPY_1144_1161 x AAA09020    .

 Alignment 1/1    1 : AAA09020  from: 1  to: 1406

          1 ValTyrGlnArg...ValLysSerGlyGlyGlnSerArgIleAl 17
            ||||||||||||||||||||||||||||||||||||||||||||||||||
    1282 ValTyrGlnArg...ValLysSerGlyGlyGlnSerArgIleAl 1299

       17 ala 20
          | |
    1300 AAGI 1303

Search completed: ...
Job time : ... secs
XX
AC   AA'54-4'; standard; DNA; 1.5 BP
XX
AC   AA'54-4';
XX
DT   28-...-... (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 2954.
XX
KW   Hybridisation assays; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathways;
KW   metabolic pathways; promoters; termination sequences; ss.
XX
OS   Arabidopsis thaliana.
XX
```



```
DE   20-SEP-1999;   99US-0154779.
DE   22-SEP-1999;   99US-0155149.
DE   23-SEP-1999;   99US-0155486.
DE   24-SEP-1999;   99US-0155659.
DE   28-SEP-1999;   99US-0156468.
DE   29-SEP-1999;   99US-0156596.
DE   04-OCT-1999;   99US-0157217.
DE   05-OCT-1999;   99US-0157754.
DE   06-OCT-1999;   99US-0157865.
DE   07-OCT-1999;   99US-0158029.
DE   08-OCT-1999;   99US-0158252.
DE   12-OCT-1999;   99US-0158869.
DE   13-OCT-1999;   99US-0159293.
DE   13-OCT-1999;   99US-0159294.
DE   13-OCT-1999;   99US-0159295.
DE   14-OCT-1999;   99US-0159329.
DE   14-OCT-1999;   99US-0159330.
DE   14-OCT-1999;   99US-0159331.
DE   14-OCT-1999;   99US-0159647.
DE   14-OCT-1999;   99US-0159648.
DE   18-OCT-1999;   99US-0159684.
DE   21-OCT-1999;   99US-0160741.
DE   21-OCT-1999;   99US-0160767.
DE   21-OCT-1999;   99US-0160768.
DE   21-OCT-1999;   99US-0160770.
DE   21-OCT-1999;   99US-0160814.
DE   21-OCT-1999;   99US-0160815.
DE   22-OCT-1999;   99US-0160980.
DE   22-OCT-1999;   99US-0160981.
DE   22-OCT-1999;   99US-0160989.
DE   25-OCT-1999;   99US-0161404.
DE   25-OCT-1999;   99US-0161405.
DE   25-OCT-1999;   99US-0161406.
DE   26-OCT-1999;   99US-0161359.
DE   26-OCT-1999;   99US-0161360.
DE   26-OCT-1999;   99US-0161361.
DE   28-OCT-1999;   99US-0161920.
DE   28-OCT-1999;   99US-0161992.
DE   28-OCT-1999;   99US-0161993.
DE   29-OCT-1999;   99US-0162142.

Alignment Scores:
              Quality:   85.00              Length:      18
                Ratio:   4.722               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 88.889

Alignment Blocks:
 US-09-817-762-2_COPY_1144_1161 x AA054645.rpt  ..

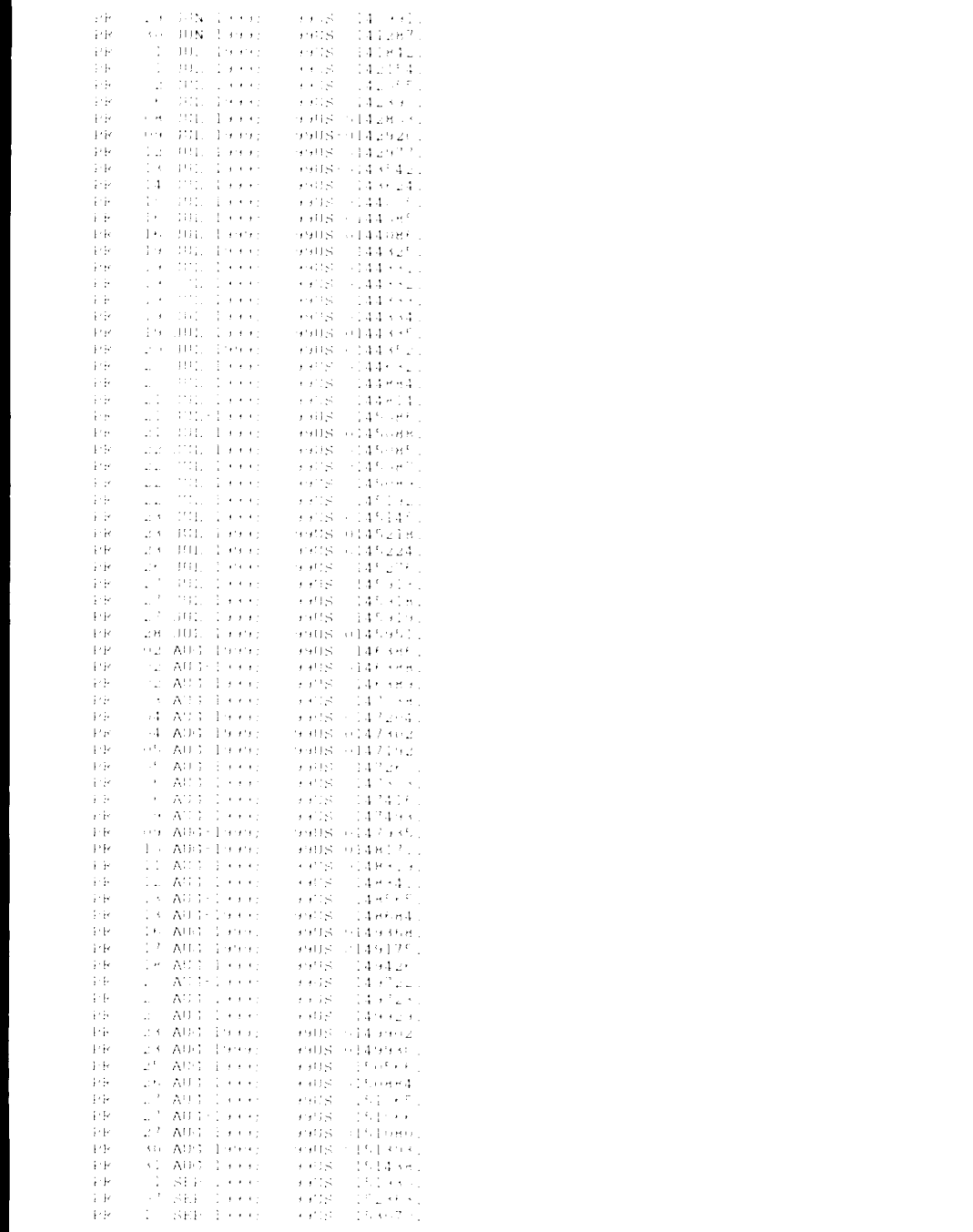
 Align set 1/1  to reverse of: AA054645  from: 1  to: 1270

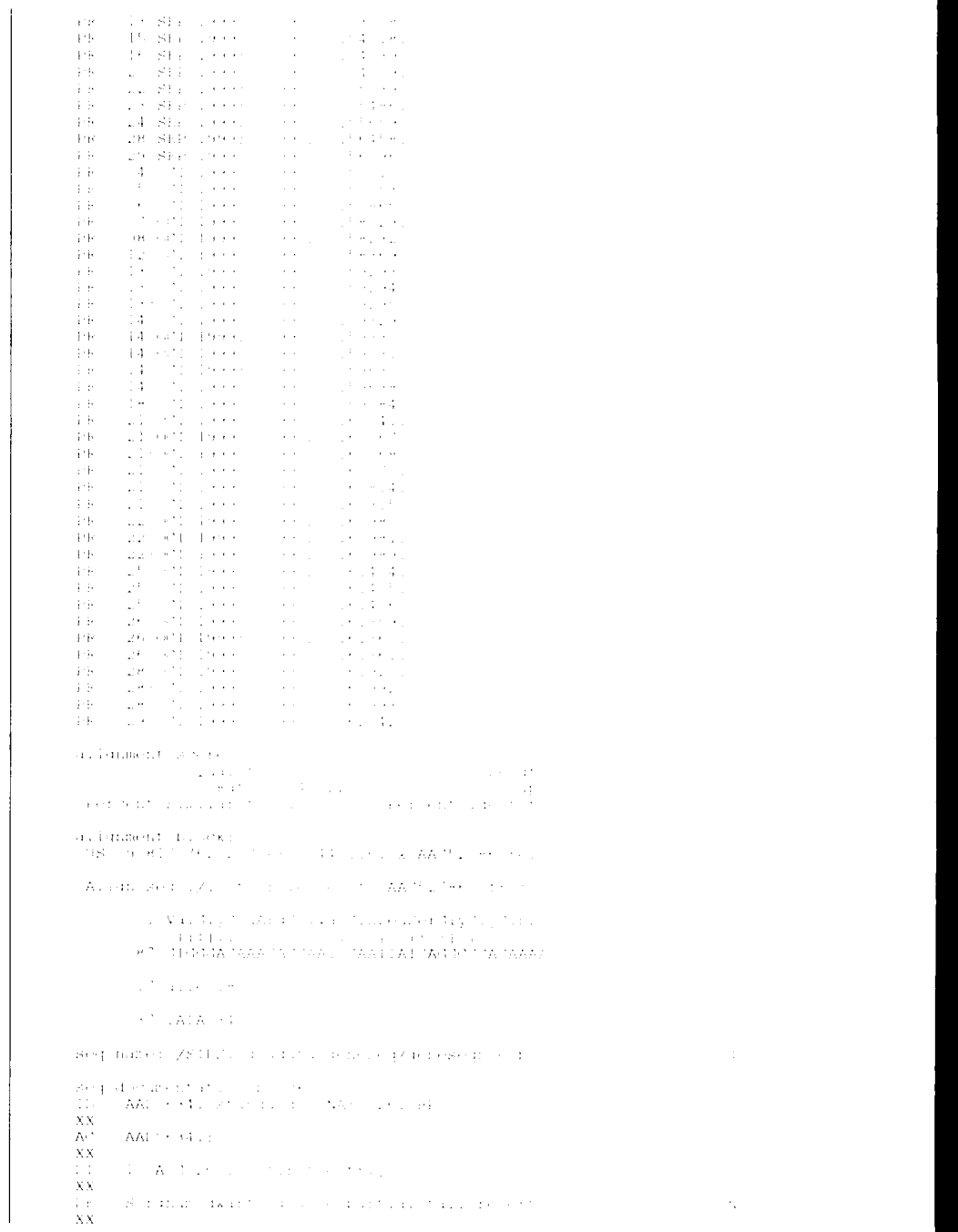
        1 ValGlyGlnArgGlyValGluLeuSerGlyGlyGlnLysGlnArgIleAl 17
          |.|:|:|||.|:|||:::|:|.|||||:|||||||.|||||||||:::|| 
       84 GlyGlyAsnAlaGlyIleAsnIleAlaLeuGlyAlaGlnValSerLeuLeu  67

       17 le 18
          |.| 
       67 leIA 51

Seq_number /????/%.!a*a/%.. .. p/%.... p. .21/???2*****.????.AA%.?%%

Seq_documentation_block:
DT   AA054645 standard; DNA; 1279 BP.
XX
AC   AA054645;
XX
DT   18-OCT-1999  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 7241.
XX
KW   hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
RN   EP1033405 A2
XX
PD   06-SEP-2000
XX
PF   25-FEB-2000; 2000EP-0301459.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123100.
PR   09-MAR-1999;   99US-0123546.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126261.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132043.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135351.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139760.
PR   18-JUN-1999;   99US-0139761.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140354.
PR   23-JUN-1999;   99US-0140394.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140821.
```





```
SW   Sorghum bicolor dwarfing gene, Dw3 protein kinase transmembrane
SW   plant resistance pathogen resistance plant breeding ds.
XX
OS   Sorghum bicolor
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..1143
FT                   /note= "Sorghum bicolor dwarfing gene,
FT                   /note= "DS does not include start codon"
FT                   /product=
XX
PN   WO2002044361-A2.
XX
PD   17-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US045176.
XX
PR   13-NOV-2000; 2000US-0247276P.
XX
PA   (GEHA/) GEN HAIL C S.
PA   (MEDI/) MEDIANI I S.
XX
PI   Bidney DL, Christou P, Lin CC, Yalpani N;
XX
DR   WPI; 2002-444299/47.
XX
PT   Novel sorghum dwarfing genes useful for modifying the growth of the
PT   organisms, particularly plants, such as maize, wheat, rice, Basmati
PT   rice, rye, millet and barley and for enhancing resistance to pathogens
PT
XX
PS   Claim 1; Page 5; 70pp; English.
XX
CC   The present sequence is partial sequence of sorghum dwarfing gene,
CC   allele, this sequence is obtained by a duplication of 882 base pairs in
CC   exon 5 of dw3 gene. Dw3 DNA is useful for modifying the growth of an
CC   organism, especially reducing the height of monocotyledon plants, such as
CC   maize, wheat, rice, Basmati rice, sorghum, rye, millet and barley. Dw3
CC   DNA is useful for producing transgenic monocotyledon and dicotyledon
CC   plants like soybean, sunflower, safflower, alfalfa, cotton, Brassica sp.,
CC   peanuts or fruit trees. The Dw3 DNA is also useful for isolating
CC   nucleotide homologous molecules that encode P-glycoproteins, enhancing
CC   the resistance of plants to pathogens, including bacteria, fungi,
CC   viruses, nematodes and insects, as probes for the isolation of other
CC   P-glycoprotein like genes and as molecular markers. Identifying plants
CC   that possess a mutant allele is useful in agriculture, particularly in
CC   breeding dwarf crop plants, particularly dwarf sorghum plants
XX
SQ   Sequence 1143 BP; 361 A; 255 C; 306 G; 221 T; 0 other;


  Alignment scores:
            Quality:   84.00             Length:      18
              Ratio:   4.667               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  94.444

 Alignment block:
 US-09-817-762-2-DY_1144-1161 x AAD06541   ..

 Alignment 1/1   to AAD06541  from: 1  to: 1261

        1 ValGlyGluArgGlyValGlnLeuSerGlyGlyGluLysGlnArgIleAl 17
          |:|:|||||:|:|||:|:|||||||||||||||||||||||||||||||||
      ... TCTTTGAATTTTTTTTGAAGTGGTGGTGAAAGTTAAGAGTGAATGT 971


  Search completed: August 19, 2002, 15:12:31
  Job time : 2111.21 secs

SeqSearch (Version 3.1.26s)
II   AAD06541 standard; DNA; 1261 BP.
XX
AC   AAD06541;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Sorghum bicolor dwarfing gene, Dw3 protein kinase transmembrane
KW   plant resistance pathogen resistance plant breeding ds.
XX
OS   Sorghum bicolor.
XX
PN   WO200144361-A2.
XX
PD   17-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US045176.
XX
PR   13-NOV-2000; 2000US-0247276P.
XX
PA   (GEHA/) GEN HAIL C S.
PA   (MEDI/) MEDIANI I S.
XX
PI   Bidney DL, Christou P, Lin CC, Yalpani N;
XX
DR   WPI; 2002-444299/47.
XX
PT   Novel sorghum dwarfing genes useful for modifying the growth of the
PT   organisms, particularly plants, such as maize, wheat, rice, Basmati
PT   rice, rye, millet and barley and for enhancing resistance to pathogens
PT
XX
PS   Claim 1; Page 5; 70pp; English.
XX
CC   The present DNA sequence is partial sequence of sorghum dwarfing gene, Dw3 mutant
CC   allele, this sequence is obtained by a duplication of 882 base pairs in
CC   exon 5 of dw3 gene. Dw3 DNA is useful for modifying the growth of an
CC   organism, especially reducing the height of monocotyledon plants, such as
CC   maize, wheat, rice, Basmati rice, sorghum, rye, millet and barley. Dw3
CC   DNA is useful for producing transgenic monocotyledon and dicotyledon
CC   plants like soybean, sunflower, safflower, alfalfa, cotton, Brassica sp.,
CC   peanuts or fruit trees. The Dw3 DNA is also useful for isolating
CC   nucleotide homologous molecules that encode P-glycoproteins, enhancing
CC   the resistance of plants to pathogens, including bacteria, fungi,
CC   viruses, nematodes and insects, as probes for the isolation of other
CC   P-glycoprotein like genes and as molecular markers. Identifying plants
CC   that possess a mutant allele is useful in agriculture, particularly in
CC   breeding dwarf crop plants, particularly dwarf sorghum plants
XX
SQ   Sequence 1261 BP; 361 A; 255 C; 306 G; 221 T; 0 other;


  Alignment scores:
            Quality:   84.00             Length:      18
              Ratio:   4.667               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  94.444

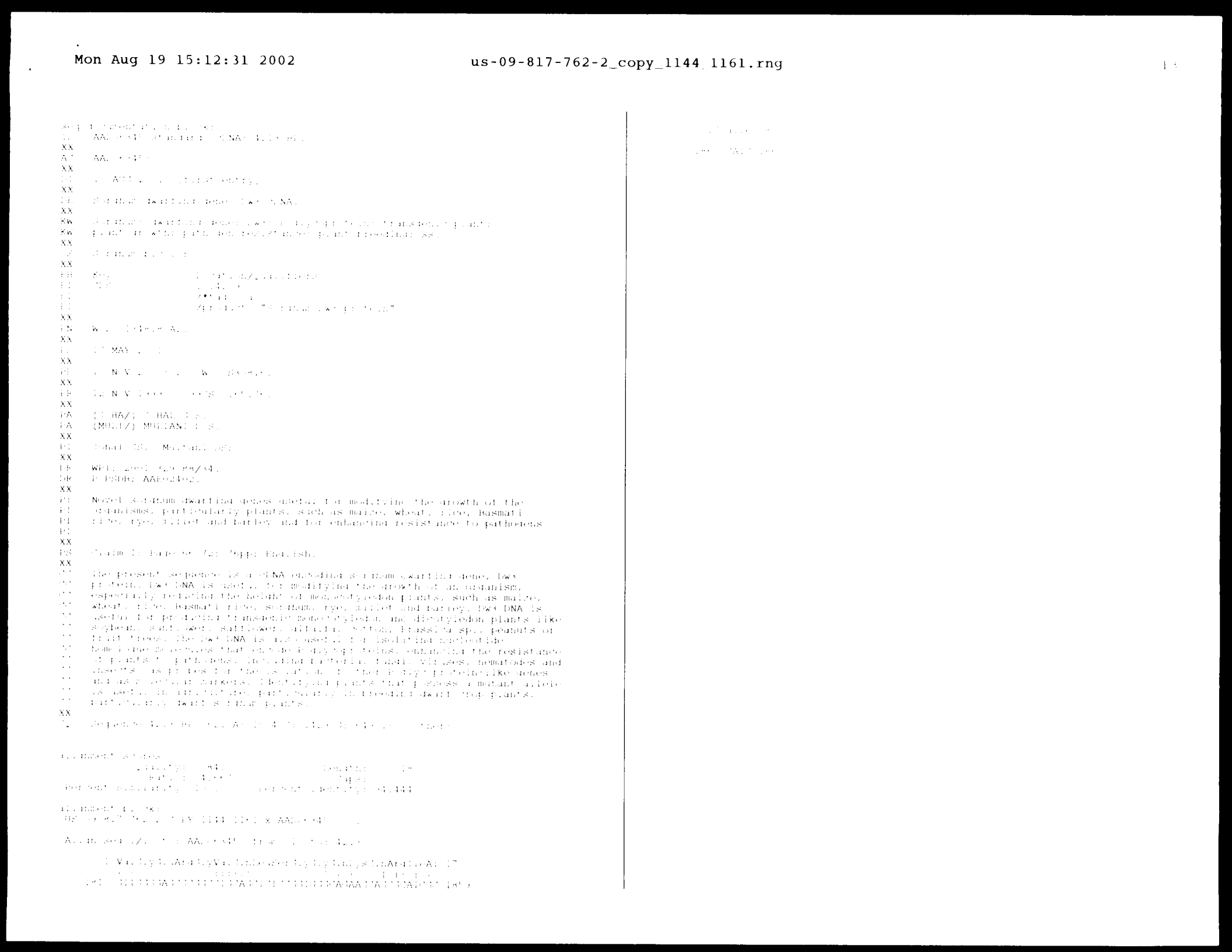
 Alignment block:
 US-09-817-762-2-DY_1144-1161 x AAD06541   ..

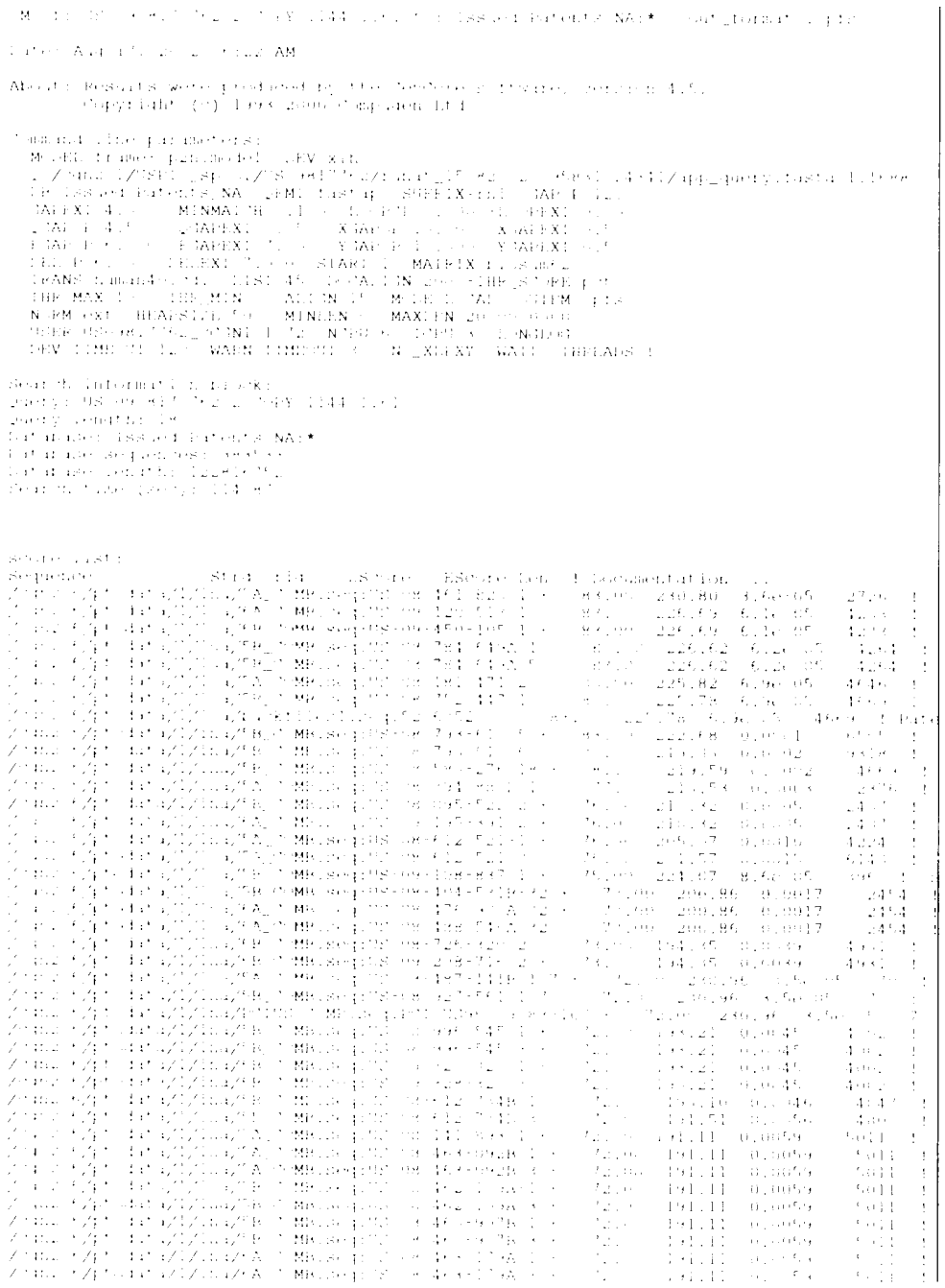
 Alignment 1/1   to AAD06541  from: 1  to: 1261

        1 ValGlyGluArgGlyValGlnLeuSerGlyGlyGluLysGlnArgIleAl 17
          |:|:|||||:|:|||:|:|||||||||||||||||||||||||||||||||
      ... TCTTTGAATTTTTTTTGAAGTGGTGGTGAAAGTTAAGAGTGAATGT 971


  Search completed: August 19, 2002, 15:12:31
  Job time : 2111.21 secs

SeqSearch (Version 3.1.26s)
```





```
            Quality:   83.00              Length:      18
              Ratio:  4.611                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 94.444

Alignment_block:
 US-09-817-762-2-COPY_1144_1161 x US-08-181-471-2 .

 Alignment 1/1   to: US-08-181-471-2  from: 1  to: 4646

        1 VaGlyGlnAspGlyValGlnAlaSerGlyGlyGlnLysGlnArgGlyAl 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1994 GlyGlnAspAsnAspAsnSerGlyGlyGlnLysGlnArgGlyAlaLe 2043

       17 aGlu 18
          |||
     2044 CAlI 2047

Seq_name: /cgn2_6/ptodata/1/1544/5E_COMB.seq:US-08-752-447-1

Seq_documentation_block:
; Sequence 1, Application US/08752447
; Patent No. 5994088
;  GENERAL INFORMATION:
;     APPLICANT:  Mechetner, Eugene
;     APPLICANT:  Robinson, Ivor B.
;     TITLE OF INVENTION:  Methods and Reagents for Preparation and
;     TITLE OF INVENTION:  Using Immunological Agents Specific for Polyoprotein
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff LLP
;       STREET:  300 South Wacker Drive, Seventh Floor
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/752,447
;       FILING DATE:  15-NOV-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mc Donnell, Thomas P.
;       REGISTRATION NUMBER:  35,808
;       REFERENCE/DOCKET NUMBER:  97,123
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312 913 0001
;       TELEFAX:  312 913 0003
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4660 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  5'UTR
;       LOCATION:  1..424
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  425..4264
;     FEATURE:
;       NAME/KEY:  3'UTR
;       LOCATION:  4265..4660
US-08-752-447-1


Alignment_scores:
            Quality:   83.00              Length:      18
              Ratio:  4.611                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 94.444

Alignment_block:
 US-09-817-762-2-COPY_1144_1161 x US-08-08-752-447-1

 Alignment 1/1   to: US-08-08-752-447-1  from: 1  to: 4669

        1 VaGlyGlnAspGlyValGlnAlaSerGlyGlyGlnLysGlnArgGlyAl 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1994 GlyGlnAspAsnAspAsnSerGlyGlyGlnLysGlnArgGlyAlaLe 2043

       17 aGlu 18
          |||
     2044 CAlI 2047

Seq_name: /cgn2_6/ptodata/1/1544/5E_COMB.seq:US-08-752-447-1

Seq_documentation_block:
;Patent No. 5206352
;     APPLICANT: Robinson, Ivor B.;Pastan Ira H.;Gottesman,
;Michael M.
;     TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
;     NUMBER OF SEQUENCES: 4
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/765,446
;       FILING DATE: 24-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 892,575
;       FILING DATE: 01-AUG-1986
;       APPLICATION NUMBER: 845,610
;       FILING DATE: 28-MAR-1986
;SEQ ID NO:3:
;       LENGTH: 4669
5206352-3


Alignment_scores:
            Quality:   83.00              Length:      18
              Ratio:  4.611                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 94.444

Alignment_block:
 US-09-817-762-2-COPY_1144_1161 x 5206352-3  ..

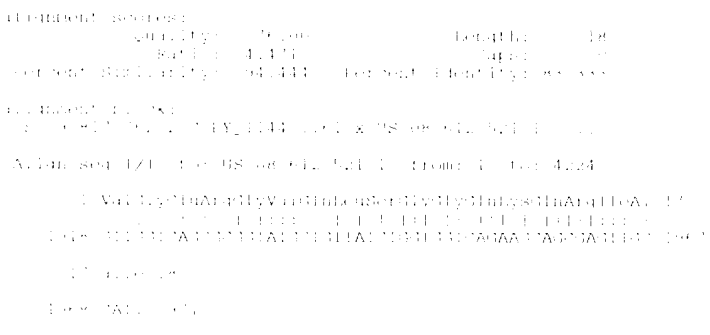
 Alignment 1/1   to: 5206352-3  from: 1  to: 4669

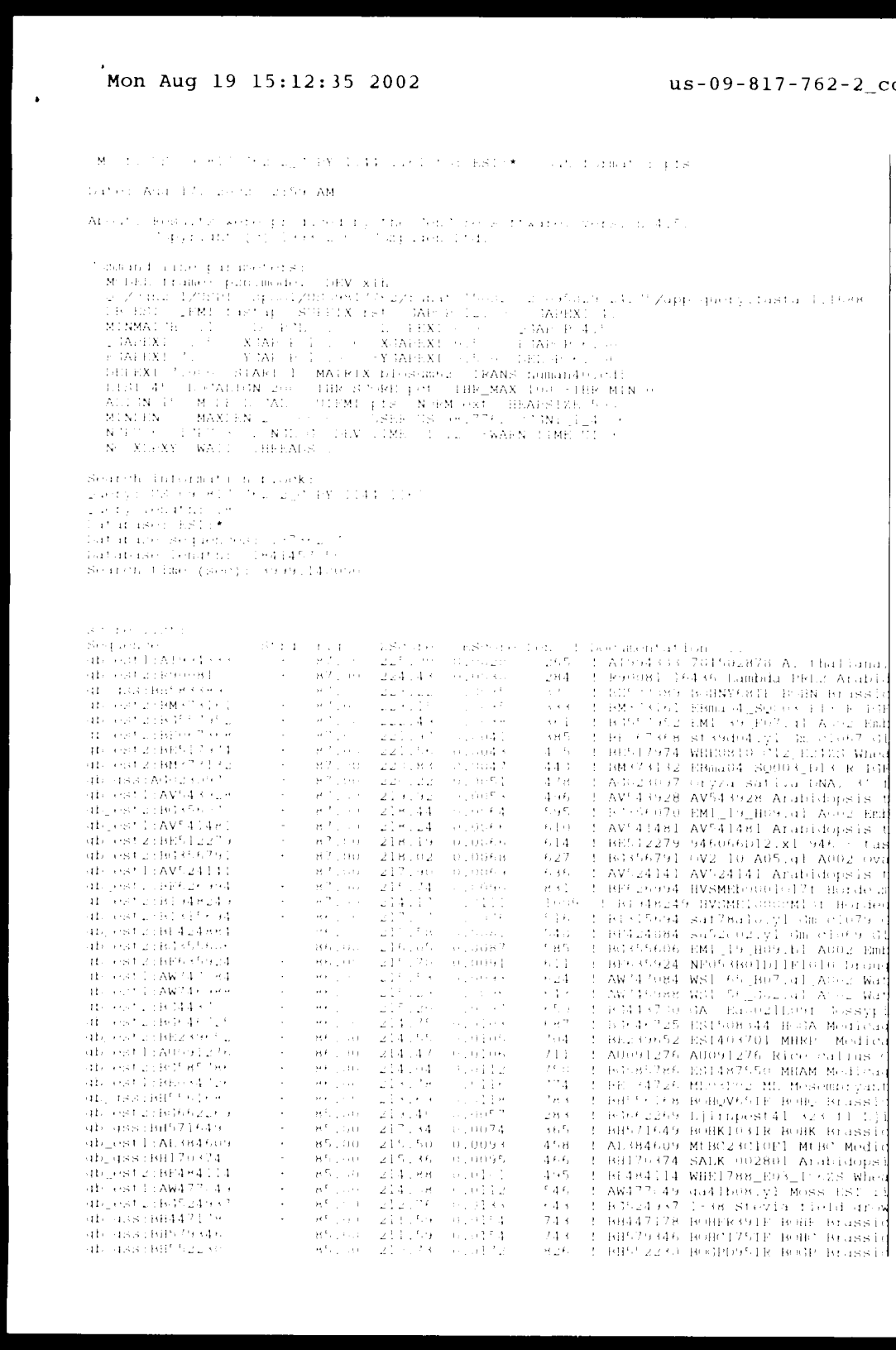
        1 VaGlyGlnAspGlyValGlnAlaSerGlyGlyGlnLysGlnArgGlyAl 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1994 GlyGlnAspAsnAspAsnSerGlyGlyGlnLysGlnArgGlyAlaLe 2043

       17 aGlu 18
          |||
     2044 CAlI 2047

Seq_name: /cgn2_6/ptodata/1/1544/5E_COMB.seq:US-08-752-447-1

Seq_documentation_block:
; Sequence 1, Application US/08/793610
; Patent No. 5858744
;  GENERAL INFORMATION:
;     APPLICANT:  BAGM, Christopher
;     APPLICANT:  ZIMMERMANN, Carol
;     APPLICANT:  STEINIAS, Wolfram
;     TITLE OF INVENTION:  RECOMBINAL VECTOR HYBRIDS AND THE USE THEREOF
;     TITLE OF INVENTION:  FOR GENE TRANSFER
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;       STREET:  655 Fifteenth Street N.W. Suite 330
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
```





```
VERSION     [illegible]
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            [illegible], Balkrishi,N., Somerville,S., Thomashow,M., Retzel
            [illegible] and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. [illegible], 1241-12[illegible] (1994)
  MEDLINE   [illegible]
COMMENT     Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE PRL  Michigan State University, East Lansing [illegible]
            Lansing MI
            Tel: [illegible]
            Fax: [illegible]
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1..284
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="[illegible]"
                     /clone_lib="Lambda PRL2"
                     /note="Vector: Lambda Zip Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA. The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques. The vector is BRL's Lambda Zip Lox. The cDNA
                     inserts were directionally cloned with Sal Not arms using
                     [illegible] primed cDNA. "
BASE COUNT     [illegible] a     [illegible] c     [illegible] g      54 t      [illegible] others
ORIGIN


alignment_scores:
            Quality:   [illegible]                Length:      18
              Ratio:   [illegible]                  Gaps:      [illegible]
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
  [illegible]

 Align seq 1/1 to: [illegible] from: 1 to: 284

        [illegible]
        [illegible]
     [illegible] AAAGAGGATTNCAGTTATCAGTTGNCAGAAATAGAGGATNGC [illegible]

        [illegible]

     [illegible]

Sequence: gb_ess:BH[illegible]

Sequence Documentation Block:
LOCUS       BH[illegible]               [illegible] bp    DNA     linear   GSS [illegible]
DEFINITION  [illegible] Brassica oleracea genomic clone [illegible], DNA
            sequence.
ACCESSION   BH[illegible]
VERSION     BH[illegible]  GI:[illegible]
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: [illegible]
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3523
            Fax: 301 838 0208
            Email: cdtown@tigr.org
            DNA is from [illegible] provided by [illegible]
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Brassica oleracea"
                     /strain="[illegible]"
                     /db_xref="taxon:3712"
                     /clone="[illegible]"
                     /clone_lib="[illegible]"
                     /note="Vector: [illegible]; Site_1: BstXI; [illegible] sheared
                     genomic DNA inserted into [illegible] using BstXI linkers"
BASE COUNT       65 a      100 c       76 g       89 t
ORIGIN


alignment_scores:
            Quality:   87.00                Length:      18
              Ratio:   4.833                  Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-817-762-2_COPY_1144_1161 x BH583885/rev  [illegible]

 Align seq 1/1 to: BH583885 from: 1 to: 330

     1 ValLeuGluArgGlyValLeuLeuSerGlyLysGlnLysGlnArgIleAl 17
       |||||||||||||||||||||||||||||||||||||||||||||||||
   205 [illegible]AGAGGATAGC 154

     [illegible]

   153 CATA 150

Sequence: gb_est:BM[illegible]

Sequence Documentation Block:
LOCUS       BM[illegible]          [illegible] bp    mRNA    linear   EST [illegible] JAN 2002
DEFINITION  EBma04 SQ00[illegible] Barley EBma04 library Hordeum vulgare cDNA
            clone EBma04 SQ00[illegible] 5', mRNA sequence.
ACCESSION   BM[illegible]
VERSION     BM[illegible]  GI:[illegible]
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; [illegible]
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   [illegible], [illegible], [illegible], McCallum,N., [illegible], [illegible]
            [illegible], Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R
            Unit of Genomics
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: +44 1382 [illegible]
```

```
                  Fax: +44 1382 562426
                  Email: [illegible]
                  All sequence has a phred quality score of 20 or over
                  Seq primer: M13 reverse.
FEATURES                   Location/Qualifiers
     source                1..333
                           /organism="Hordeum vulgare"
                           /cultivar="Optic"
                           /db_xref="taxon:4513"
                           /clone="EBma04_SQ003_F13"
                           /clone_lib="IGF Barley EBma04 library"
                           /tissue_type="Maternal tissue"
                           /dev_stage="10 days post anthesis"
                           /lab_host="DH10B"
                           /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                           Non normalised library, directionally cloned into pSPORT1.
                           Derived from maternal tissue dissected from developing
                           grains (10 days post anthesis) in glasshouse grown barley
                           plants. Developed as part of the barley transcriptome
                           resources of BBSRC/SEERAD funded cereal IGF (Investigating
                           Gene Function) project."
BASE COUNT     [illegible] a   [illegible] c   12[illegible] g   44 t
ORIGIN

Alignment Scores:
          Quality:     87.00              Length:      18
            Ratio:     4.833                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

Alignment Block:
 US-09-817-762-2_COPY_1144_1161 x BM373161 ..

 Align seq 1/1 to: BM373161 from: 1 to: 333

        1 ValGlyGluArgGlyValGlnLeuSerGlyGlyGlnLysGlnArgIleAl 17
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       65 GTCGGGGA[illegible]GGCGTGCAGCTGTCGGGCGGCCAGAAGCAGCGTATCGC 114

       17 aIle 18
          ||||
      115 CATT 118

Seq name: gb_est2:BG557952

Seq Documentation Block:
LOCUS       BG557952               361 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  EM1_50_F07.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG557952
VERSION     BG557952.1  GI:13580[illegible]
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Reid,S.P., Cordonnier Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 4[illegible]
```

```
FEATURES             [illegible]
     source          [illegible]
BASE COUNT           [illegible]
ORIGIN

Alignment Scores:
 [illegible]
 Percent Similarity: [illegible]    Percent Identity: [illegible]

Alignment Block:
 US-09-817-762-2_COPY_1144_1161 x [illegible]

 Align seq 1/1 to: [illegible]

 [illegible]

Seq name: [illegible]

Seq Documentation Block:
LOCUS       [illegible]
DEFINITION  [illegible]
            mRNA sequence.
ACCESSION   [illegible]
VERSION     [illegible]
KEYWORDS    EST.
SOURCE      [illegible]
  ORGANISM  [illegible]
REFERENCE   1  [illegible]
  AUTHORS   [illegible]
  TITLE     [illegible]
  JOURNAL   Unpublished [illegible]
COMMENT     [illegible]
FEATURES             Location/Qualifiers
     source          [illegible]
```

```
                     /clone_lib="Gmelo06J"
                     /tissue_type="Germinating shoots, 3 day old seedling, grown
                     in vermiculite"
                     /cultivar="Gladiator"
                     /db_xref="taxon:4565"
                     /note="Vector: pBluescript II SK(+); Site 1: EcoRI site at
                     XhoI: The cDNA library was constructed from mRNA isolated
                     from germinating shoots of 3 day old seedling for the
                     cultivar Williams 82. The seedlings were germinated in a
                     growth chamber (Sheldon germination paper in a solution
                     containing 1mM gibberellic acid); complementary DNA was synthesized
                     from mRNA using a primer consisting of a poly(dT) sequence
                     with a XhoI restriction site; EcoRI adapters were ligated
                     to the blunt ended cDNA fragments followed by XhoI
                     digestion. The cDNA fragments were directionally cloned
                     into the EcoRI-XhoI restriction site of the pBluescript
                     SK(+). The library cDNA fragments were transformed into
                     DH5 host cells (Stratagene). This library was constructed
                     in the laboratory of Dr. Randy Shoemaker."
BASE COUNT        xxx a    xxx c    xxx g    xxx t
ORIGIN


Alignment Scores:
            Quality:   87.00              Length:      18
              Ratio:  4.833                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment 1, BE:
US-09-817-762-2 (1144_1161 x BE517974    ..

 Alignment 1/1  for BE517974  from: 1  to: 485

        Qy  Va.Cy.TuAr.HiGlyVa.HiLeuSe.GlGlyGlGlIlLysGlnArgLeuAl 17
            : : : : : : : : : : : : : : : : : : : : : : : : : : :
        Db  TTTTTGAGACGCGTG AATCATGCAAAA AAAGGATCC TTG

      17 aller 18

      94 "AT" 97

Sequence BE517974

Sequence Characteristics:

Sequence Documentation Block:
LOCUS       BM477962                 485 bp    mRNA    linear   EST 10-JAN-2002
DEFINITION  Edward_S_et_al_Lys_E_EST_Barley_EMmar4_library_BE_clone_Elmar4_cDNA
            clone Elmar4_c2074_1_F1 mRNA sequence
ACCESSION   BM477962
VERSION     BM477962.1  GI:18126922
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Hedley,C.L., Lundqvist,U., Franckowiak,J., McMaster,N., Matthews,D.,
            Ramsay,L., Marshall,D.F., Waugh,R. and Wanamaker,S.
  TITLE     Development of Barley Transcript Map Resources
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 485)
  AUTHORS   Waugh,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Plant Genomics, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: +44 1382 562731
            Fax: +44 1382 562426
            Email: rwaugh@scri.sari.ac.uk
            All sequences has a Phred quality score of 15 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1..485
                     /organism="Hordeum vulgare"
                     /organism "triticum aestivum"
                     /cultivar "Chinese Spring"
                     /db_xref "taxon:4565"
                     /clone="Wbl0001_12_L4"
                     /note="Wheat seedling 5 cm shoot cDNA library"
                     /tissue_type="Crown tissue of seedling"
                     /dev_stage="Five-week old seedling"
                     /lab_host="E. coli SURE"
                     /note="Vector: Lambda ZAP Xho; excised phagemid
                     size: 3.0kb; site 1: Xho; site 2: Xho; Stratagene cDNA
                     grown at 4 C for 5 weeks.  The tissue, total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA clones were in vivo excised to give pBluescript
                     plasmids in the E. coli host (Short, Close, Fonstad) of
                     the University of California, Riverside, Fossil DNA
                     preparation and DNA sequencing were performed at the J.
                     Anderson Lab (et. Other authors)."
BASE COUNT        xxx a    xxx c    xxx g    xxx t
ORIGIN


Alignment Scores:
            Quality:   87.00              Length:      18
              Ratio:  4.833                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment 1/1:
US-09-817-762-2 (1144_1161 x BE517974)    ..

 Alignment 1/1  for BE517974  from: 1  to: 485

        Qy  Va.Cy.TuArgHiGlyVaHiLeuSerGlyGlyGlnLysGlnArgLeuAl 17
            : : : : : : : : : : : : : : : : : : : : : : : : : :
        Db  TTTTTGAGACGCGTGAATCATGCAAAAAAAGGATCCTTG

      17 aller 18

      94 "AT" 97

Sequence BE517974

Sequence Characteristics:

Sequence Documentation Block:
LOCUS       BM477962                 485 bp    mRNA    linear   EST 10-JAN-2002
DEFINITION  Edward_S_et_al_Lys_E_EST_Barley_EMmar4_library_BE_clone_Elmar4_cDNA
            clone Elmar4_c2074_1_F1 mRNA sequence
ACCESSION   BM477962
VERSION     BM477962.1  GI:18126922
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Hedley,C.L., Lundqvist,U., Franckowiak,J., McMaster,N., Matthews,D.,
            Ramsay,L., Marshall,D.F., Waugh,R. and Wanamaker,S.
  TITLE     Development of Barley Transcript Map Resources
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 485)
  AUTHORS   Waugh,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Plant Genomics, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: +44 1382 562731
            Fax: +44 1382 562426
            Email: rwaugh@scri.sari.ac.uk
            All sequences has a Phred quality score of 15 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1..485
                     /organism="Hordeum vulgare"
```

```
                     /cultivar "Nipponbare"
                     /db_xref "taxon:4530"
                     /clone "EBMand_S2003 GLS"
                     /clone_lib "IGF Barley EBMand library"
                     /tissue_type "Maternal tissue"
                     /dev_stage "10 days post anthesis"
                     /lab_host "DH10B"
                     /note "Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1,
                     derived from maternal tissue dissected from developing
                     grains (10 days post anthesis) in glasshouse grown barley
                     plants. Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded retreat IGF (Investigating
                     Gene Function) project."
BASE COUNT      [illegible]
ORIGIN


alignment scores:
            Quality:  0.000              Length:      0
              Ratio:  4.833                Gaps:      1
 Percent Similarity: [illegible]  Percent Identity: [illegible]

alignment_block:
 US-09-817-762-2 (1144_1161 x BM837182  ..

 Align Seq I/1  1 to BM837182  from 1  to 446

        1 ValGlyGluAspGlyValGlnGlnSerGlyGlyGlnLysGluArgGlyAl 17
          ||||| |||||| |||||||||||||||||||||||| ||||||||||||
      447 GluAGGAGAAGGCGGTCAGCAGTCGCGGGGGCAGAAGGAGAGGGAGGC 496

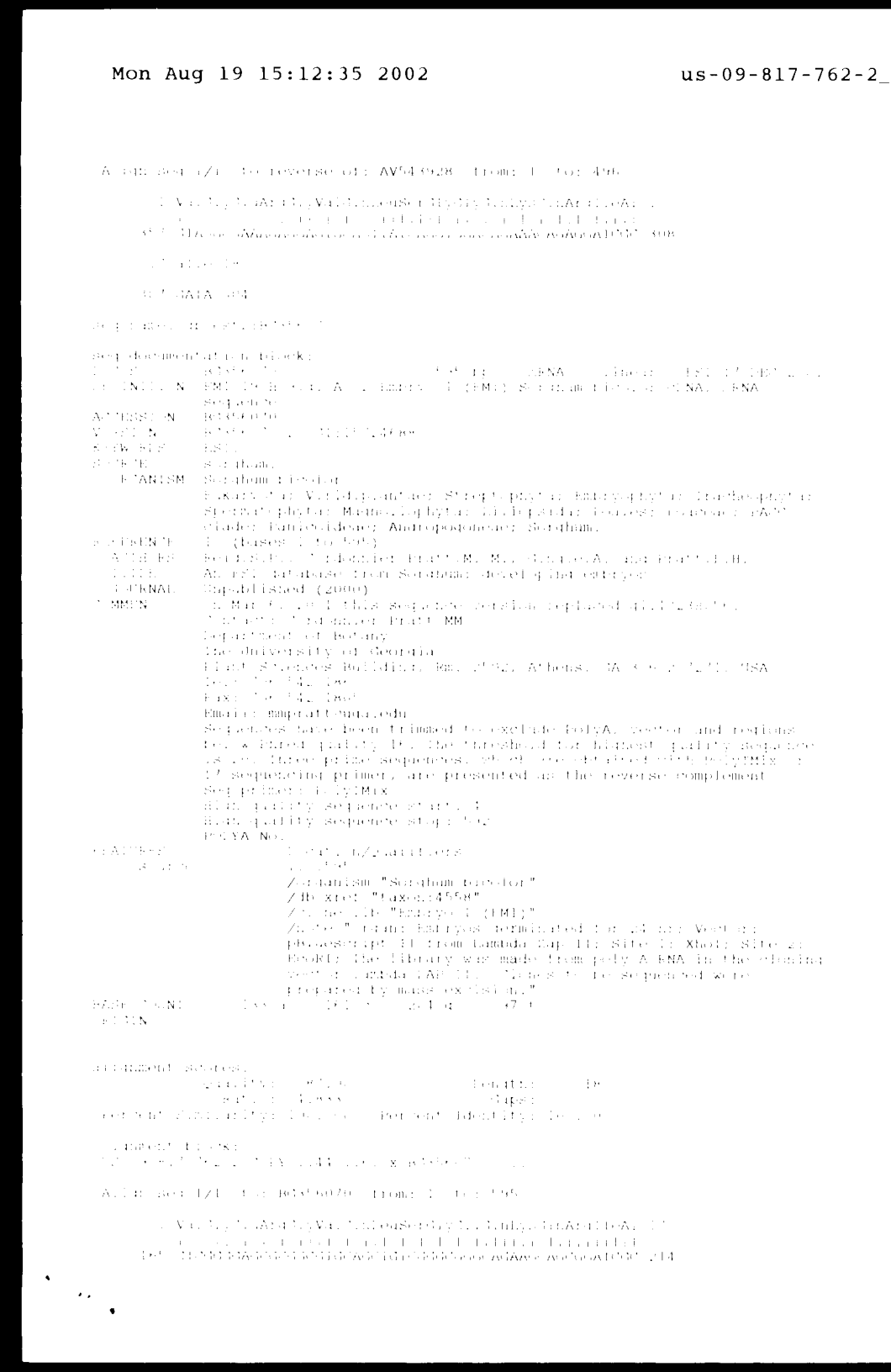
       17 alGlu 18
          |||
      497 GAG 519

Seq number: Query: 2; Type:

seq_documentation_block:
LOCUS       AV[illegible]                          [illegible] bp
DEFINITION  AV[illegible] [illegible]Hordeum vulgare [illegible]
            mRNA [illegible]
ACCESSION   AV[illegible]
VERSION     AV[illegible]
KEYWORDS    EST.
SOURCE      Hordeum [illegible]
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Hordeum.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Asakura,Y., Nagamura,Y., Sato,K. and
  TITLE     A survey of full-length sequences of cDNA in Al
            of the [illegible] full-length expressed [illegible]
            size [illegible] cDNA libraries
  JOURNAL   DNA Res. [illegible] (in press)
  MEDLINE   [illegible]
COMMENT     Contact:
            The full length cDNA clone [illegible]
            Kazuki Saito, Plant Genome [illegible]
            Yama[illegible], Tsukuba, Ibaraki [illegible]
            Email: [illegible]
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism "Hordeum vulgare"
                     /cultivar "[illegible]"
                     /db_xref "taxon:[illegible]"
                     /clone "[illegible]"
                     /clone_lib "[illegible]"
                     /tissue_type "[illegible]"
                     /dev_stage "[illegible]"
                     /lab_host "[illegible]"
                     /note "Vector: [illegible]"
BASE COUNT      [illegible]
ORIGIN


alignment scores:
            Quality:  [illegible]              Length:      [illegible]
              Ratio:  [illegible]                Gaps:      [illegible]
 Percent Similarity:  [illegible]  Percent Identity:  [illegible]

alignment_block:
 US-09-817-762-2 (1144_1161 x AV[illegible]
```



```
                Eukaryotae; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
                clade; Panicoideae; Andropogoneae; Zea.
  REFERENCE     1  (bases 1 to 614)
    AUTHORS     Walbot,V.
    TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
                University
    JOURNAL     Unpublished (1999)
  COMMENT       Contact: Walbot V
                Department of Biological Sciences
                Stanford University
                 Accession AI491004.1 GI:4
                Fax: +1 650 723 6132
                Email: walbot@stanford.edu
                Clones sequenced from 5' to 3'
  FEATURES             Location/Qualifiers
       source          1..614
                       /organism="Zea mays"
                       /cultivar="B84"
                       /db_xref="taxon:4577"
                       /clone_lib="cg - tassel primordium prepared by Schmidt
                       Lab"
                       /tissue_type="tassels"
                       /dev_stage="1st after the transition from vegetative to
                       reproductive development"
                       /lab_host="XL1-B"
                       /note="Organ: tassels; Vector: Uni-ZAP; Site_1: EcoRI;
                       Site_2: XhoI; George Chuck dissected immature tassels
                       between 1mm and 3mm. Sharon Stanfield prepared the cDNA
                       library in Uni-ZAP. Sample insert size range was 450 bp
                       to 3 Kb with a 1 Kb average."
  BASE COUNT        114 a    205 c    165 g    130 t
  ORIGIN


alignment scores:
                Quality:   82.00            Length:      14
                  Ratio:    1.833             Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
US-09-817-762-2_copy_1144_1161.rst x BE512279/rc

Align Seq 1/1   1 reverse of: BE512279  from:  1  to: 614

        1 VarSeqGlnArgHysValGlnIleSerGlyGlyHisLysGlnArgHisA 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      614 GTCGCGCAGTCCTCCTCCACCTCCTCCGCGTCGCAGAAGCACGCGCATG 567

       17 aHisTrp 18
          |  |
      566 GATGCAC 544


Seq_Number: BE512279.1 ;

Seq_documentation_block:
LOCUS       BE512279                 614 bp    mRNA    linear   EST 06-MAR-2001
DEFINITION  cVZ_1c_A_1_1_A06_2_vary_a (cVZ) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE512279
VERSION     BE512279.1  GI:7242877
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryotae; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Sreedasyam,A., Mian,R., Gingle,A., Mensalia,L., Sudmant,M. and Pratt
            L.H.
  TITLE     AG EST literature from Sorghum varieties of varying immature stages
  JOURNAL   Unpublished (2000)
  COMMENT   Contact:  Sreedasyam A.
            Plant Biology Division, Samuel Roberts Noble Foundation
            Ardmore, OK 73401
            Tel: 580-224-6600
            Fax: 580-224-6692
            Email: asreedasyam@noble.org
            Clones sequenced from 5' to 3'
  FEATURES             Location/Qualifiers
       source          1..614
                       /organism="Sorghum bicolor"
                       /db_xref="taxon:4558"
                       /cultivar="BTx623"
                       /dev_stage="Stage 2 (V2)"
                       /clone="cVZ_1c_A_1_1_A06_2_vary_a"
                       /lab_host="DH10B"
                       /clone_lib="cVZ_1c_A_1_1 Sorghum bicolor cDNA, mRNA
                       sequence"
                       /tissue_type="Apical meristem"
                       /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                       XhoI; Sreedasyam A. Sorghum bicolor cDNA, mRNA sequence"
  BASE COUNT        114 a    205 c    165 g    130 t
  ORIGIN


alignment scores:
                Quality:   82.00            Length:      14
                  Ratio:    1.833             Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
US-09-817-762-2_copy_1144_1161.rst x BE512279/rc

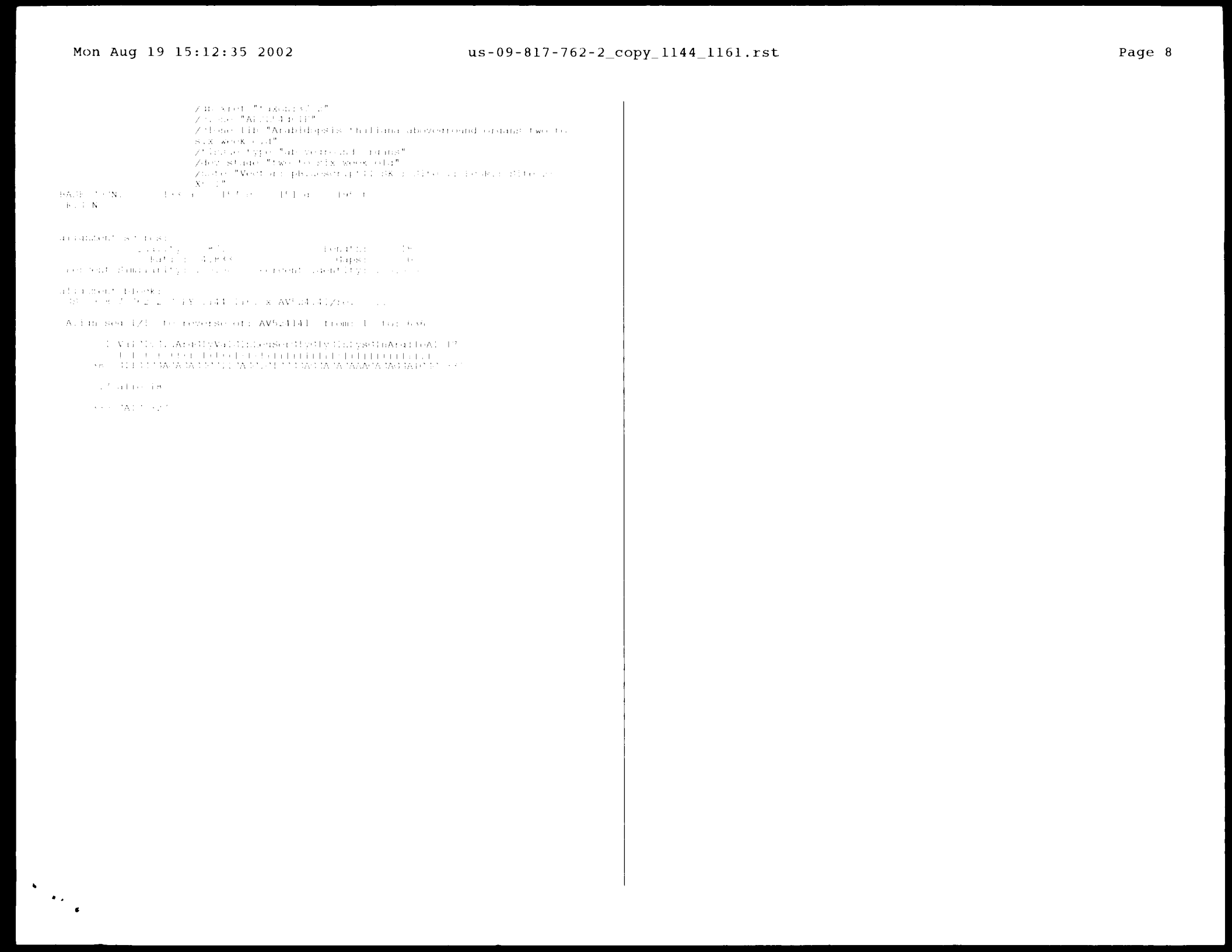
Align Seq 1/1   1 reverse of: BE512279  from:  1  to: 614

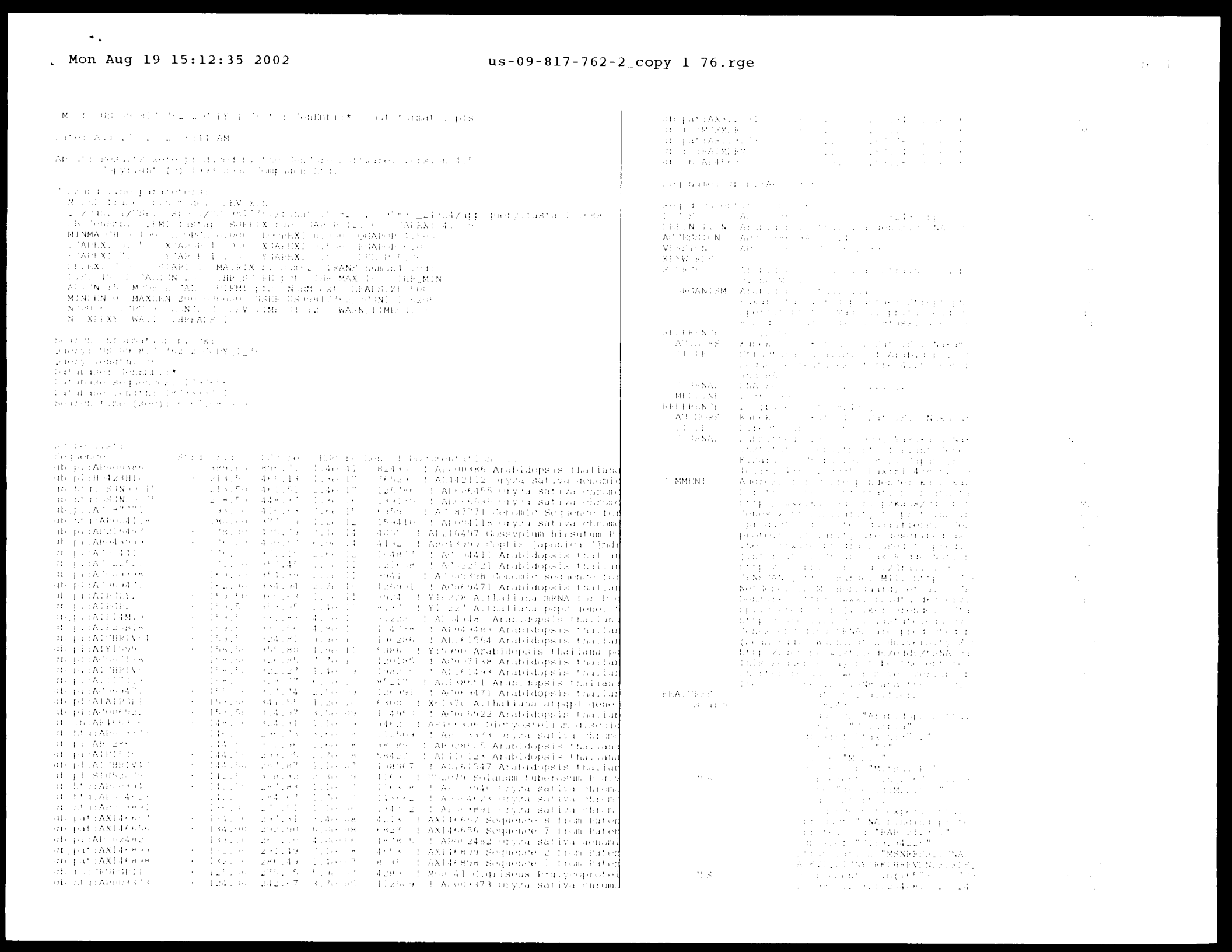
        1 VarSeqGlnArgHysValGlnIleSerGlyGlyHisLysGlnArgHisA 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      614 GTCGCGCAGTCCTCCTCCACCTCCTCCGCGTCGCAGAAGCACGCGCATG 567

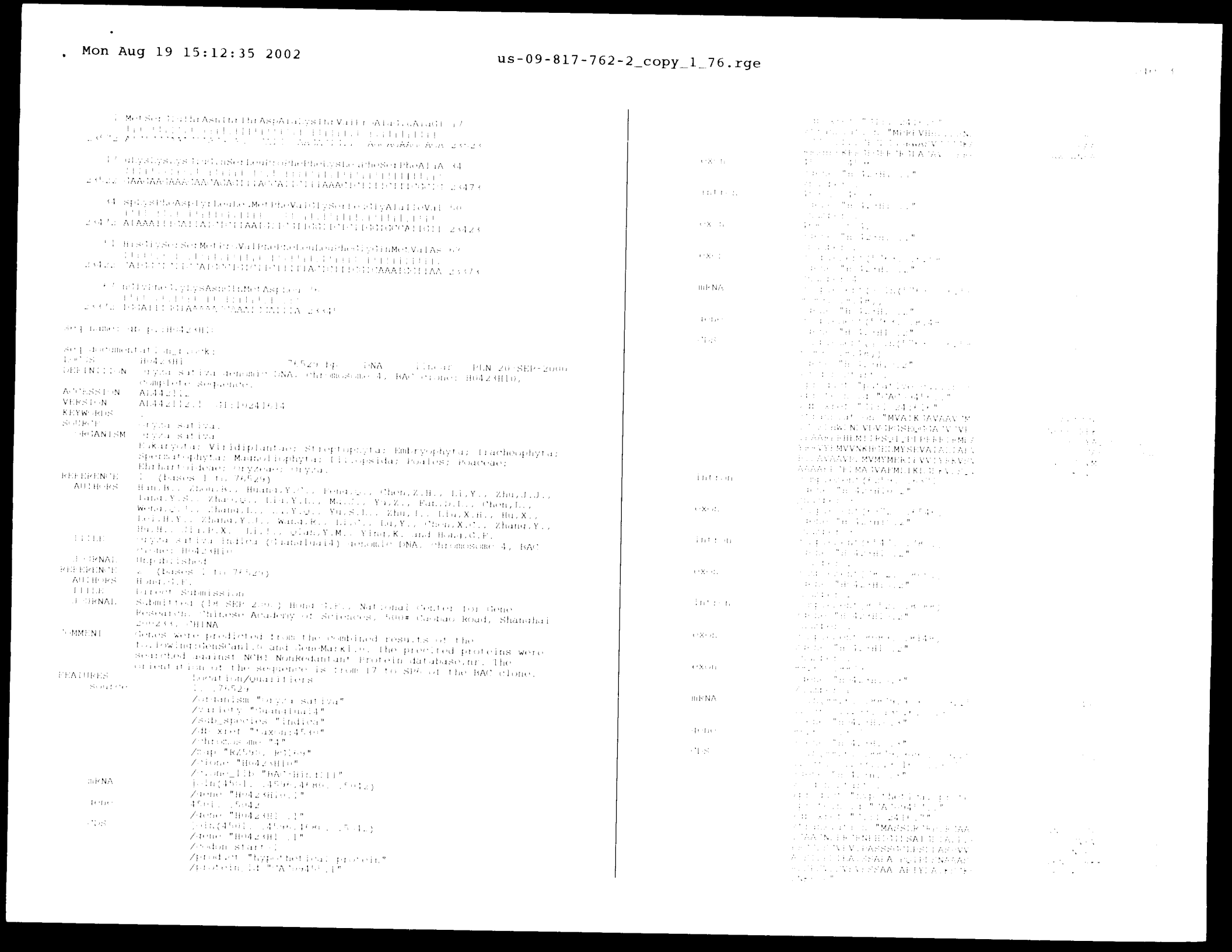
       17 aHisTrp 18
          |  |
      566 GATGCAC 544


Seq_Number: BE512279.1 ;

Seq_documentation_block:
LOCUS       AV524131                 614 bp    mRNA    linear   EST 06-MAR-2001
DEFINITION  AV524131 Arabidopsis thaliana cDNA clone RAFL05-
            -21-A21 5', mRNA sequence.
ACCESSION   AV524131
VERSION     AV524131.1  GI:7242878
KEYWORDS    EST.
SOURCE      Arabidopsis
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Asamizu,E., Nakajima,M., Kaneko,T., Miyajima,N., Sato,S., Kotani,H.,
            Tabata,S. and Kato,T.
  TITLE     A large scale structural analysis of cDNAs in Arabidopsis
            thaliana
  JOURNAL   DNA Res. 7 (2), 175-180 (2000)
  MEDLINE   20341477
  COMMENT   Contact:  Asamizu E.
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp
  FEATURES             Location/Qualifiers
       source          1..614
                       /organism="Arabidopsis thaliana"
                       /db_xref="taxon:3702"
                       /clone="RAFL05-21-A21"
                       /lab_host="DH10B"
                       /clone_lib="Arabidopsis thaliana
```







LOCUS AL442112 76529 bp DNA linear PLN 20-SEP-2000
DEFINITION Oryza sativa genomic DNA, chromosome 4, BAC clone: B0423H10, complete sequence.
ACCESSION AL442112
VERSION AL442112.1 GI:19241914
KEYWORDS .
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 76529)
AUTHORS Rao,B., Zhou,B., Huang,Y.C., Feng,Z., Chen,Z.B., Li,Y., Zhu,J.J., Lin,Y.S., Zhang,Z., Li,Y.L., Ma,Z., Yu,Z., Fan,D.L., Chen,L., Wei,Z., Zhang,L., Yu,Y., Yu,S.L., Zhu,L., Liu,X.H., Hu,X., Lei,H.Y., Zhang,Y.L., Wang,R., Li,C., Lu,Y., Chen,X.C., Zhang,Y., Hu,H., Zha,P.X., Li,L., Qian,Y.M., Ying,K. and Hong,G.F.
TITLE Oryza sativa Indica (Guangluai4) genomic DNA, chromosome 4, BAC clone: B0423H10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 76529)
AUTHORS Hong,G.F.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-2000) Hong G.F., National Center for Gene Research, Chinese Academy of Sciences, 500# Caobao Road, Shanghai 200233, CHINA
COMMENT Genes were predicted from the combined results of the following programs GeneScan1.0 and GeneMark1.0. The predicted proteins were searched against NCBI NonRedundant Protein database. The orientation of the sequence is from 17 to 5P6 of the BAC clone.

FEATURES Location/Qualifiers
source 1..76529
/organism="Oryza sativa"
/variety="Guangluai4"
/sub_species="indica"
/db_xref="taxon:4530"
/chromosome="4"
/map="KZ59-, p22.3"
/clone="B0423H10"
/clone_lib="BAC-HindIII"
mRNA join(45..1459,4486..4560)
/gene="B0423H10.1"
gene 45..4560
/gene="B0423H.1"
CDS join(45..1459,4486..4560)
/gene="B0423H.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG94560.1"

```
intron          [illegible]..19089
                /gene "H0423H10.3"
                /number 1
exon            [illegible]
                /gene "H0423H10.3"
                /number [illegible]
intron          [illegible]
                /gene "H0423H10.3"
                /number 2
exon            [illegible]
                /gene "H0423H10.3"
                /number [illegible]
intron          [illegible]
                /gene "H0423H10.3"
                /number [illegible]
exon            [illegible]
                /gene "H0423H10.3"
                /number 4
intron          [illegible]
                /gene "H0423H10.3"
                /number [illegible]
exon            [illegible]..12931
                /gene "H0423H10.3"
                /number 5
intron          [illegible]
                /gene "H0423H10.3"
                /number [illegible]
exon            [illegible]..[illegible]
                /gene "H0423H10.3"
                /number [illegible]
exon            complement([illegible])
                /gene "H0423H10.4"
                /number 2
mRNA            complement(join([illegible]))
                /gene "H0423H10.4"
gene            complement([illegible])
                /gene "H0423H10.4"
CDS             complement(join([illegible]))
                /gene "H0423H10.4"
                /codon_start 1
                /product "hypothetical protein"
                /protein_id "CAC09458.1"
                /db_xref "GI:10241618"
                /translation "MEMWKLRQEKRLLAAPISRSIPKLYPVP[illegible]
                PGSRSA[illegible]
                [illegible]LLSSLLHRRRPALLLSGFIRVPIGRRNKAPA[illegible]AIVALLIAMMPVLAPV
                LWRVEGVPKL[illegible]GKIKGRP[illegible]RVLLQLGP[illegible]"
intron          complement([illegible])
                /gene "H0423H10.4"
                /number 1
exon            complement([illegible])
                /gene "H0423H10.4"
                /number 1
mRNA            join(20660..20812,20995..21500,22319..22324)
                /gene "H0423H10.5"
gene            20660..22324
                /gene "H0423H10.5"
CDS             join(20660..20812,20995..21500,22319..22324)
                /gene "H0423H10.5"
                /codon_start 1
                /product "hypothetical protein"
                /protein_id "CAC09459.1"
                /db_xref "GI:10241619"
                /translation "MAM[illegible]
                YEG[illegible]SAA[illegible]INSA[illegible]VSALF[illegible]AASALP
                ASIASPRSCVPGFPNIQSAFFNIRLSLLIWILKLVQERFSGLLSLLSSLLVLLIGGWQ
                IAYLKEAVEKK[illegible]AAVA[illegible]KAVKA[illegible]SHLIKKPF
                KLLSHLVKLQSEKLLEPREAEAKSQAKEVKLQEKS"
exon            20660..20812
                /gene "H0423H10.5"
                /number 1
intron          20813..20994
                /gene "H0423H10.5"
                /number 1
exon            20995..21500
                /gene "H0423H10.5"
                /number 2
intron          21501..22318

alignment_scores:
           Quality:  [illegible]          Length:      7[illegible]
             Ratio:  [illegible]            Gaps:       [illegible]
 Percent Similarity: [illegible]   Percent Identity: [illegible]

alignment_block:
 [illegible] x H0423H10 [illegible]

 Align seq 1/1 to: H0423H10 from: 1 to: [illegible]

      [illegible] AsnThrThrAspAlaLys[illegible]ValPro[illegible]
          |||       |||||           ::::|::::::::||:::   ::
  [illegible] AA[illegible]AAGAA[illegible] 32447

     [illegible] [illegible]SerLeu[illegible]PheLysLeuPheSerPheAlaAspLysPheAsp [illegible]
         [illegible]
  [illegible] 32[illegible]

     [illegible] [illegible]LeuMetPheVal[illegible]SerLeu[illegible]AlaIleValHis[illegible]SerSer 54
         [illegible]
  [illegible] 32[illegible]

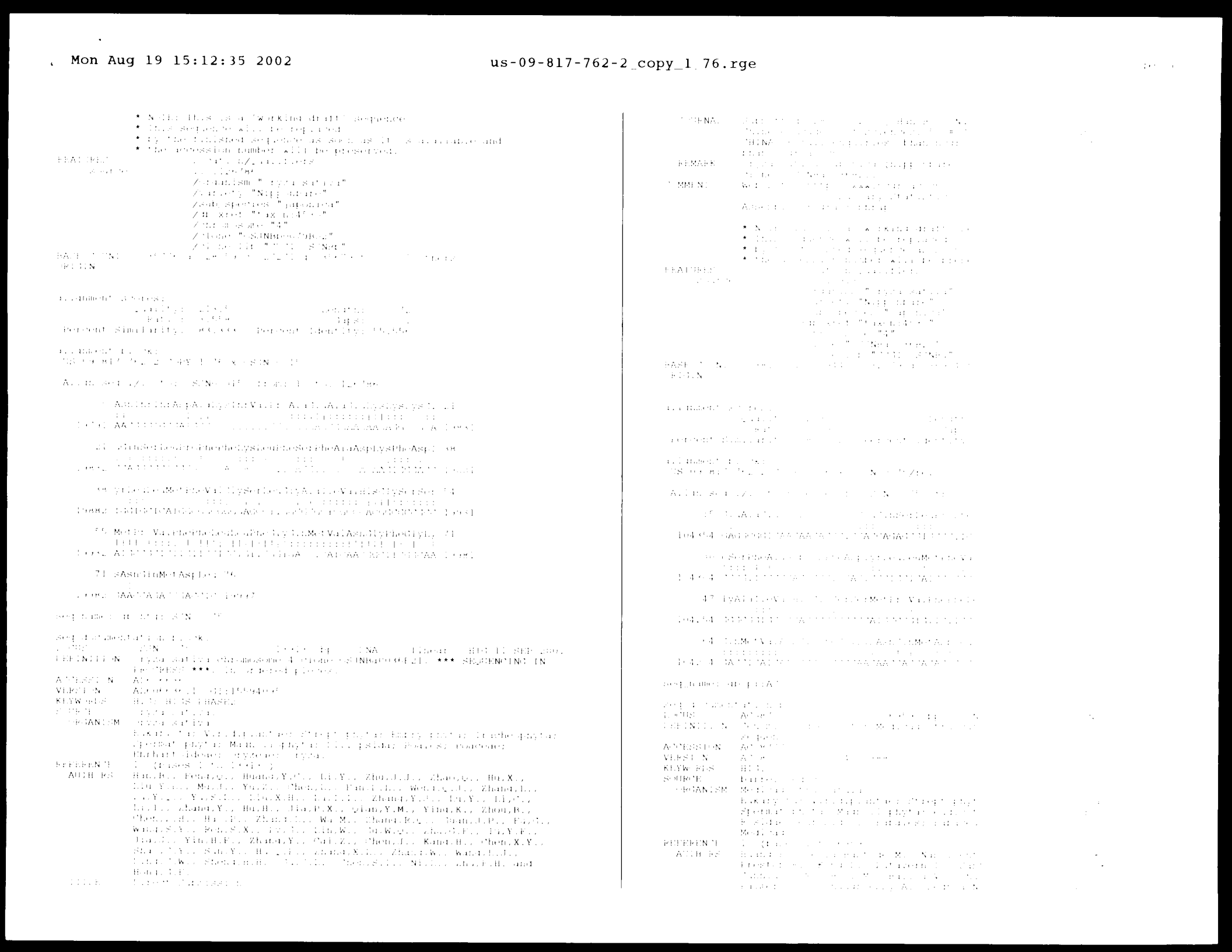
     [illegible] Met[illegible]ValPhePhe[illegible]LeuPheGlyGlnMetValAsnGlyPhe[illegible] 71
         [illegible]
  [illegible] 32[illegible]

     71 [illegible]AsnGlnMetAspLeu 76
         [illegible]
  [illegible] 32626

seq_name: gb_htg:OSJN00015

seq_documentation_block:
LOCUS       OSJN00015          126786 bp    DNA     linear   HTG [illegible]
DEFINITION  Oryza sativa chromosome 4 clone OSJNBb0079B02, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL606455
VERSION     AL606455.1  GI:[illegible]
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 126786)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            [illegible], Mu,J., Y[illegible], Chen,[illegible], Fan,[illegible], Weng,[illegible], Zhang,[illegible],
            Lu,Y.[illegible], Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.L., Lu,Y., Li,C.,
            [illegible], Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zh[illegible],
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,K.[illegible], Guan,J.F., Fu,[illegible],
            W[illegible],S.Y., Ren,S.X., [illegible], Lin,W., [illegible],W.[illegible], Zhu,G.F., [illegible],Y.F.,
            Tia[illegible], Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
            Shi,[illegible]Y., Sun,Y., Hu[illegible],P., Zhang,X.L., Zhang,W., Wang,[illegible],
            [illegible],C.W., Sheng,H.H., Gu,J.L., Chen,S.L., Ni,L., Zhu,F.H. and
            Hong,[illegible]
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2000) Han Bin, National Center for Gene Research,
            Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai [illegible],
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
REMARK      Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBb0079B02.
COMMENT     Web site: http://[illegible]
                              Summary Statistics
            Assembly program: phrap

            This is a complete sequence.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (20 AUG 2001) Rika Annotation Group, Genome
            Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown
            Road, Cold Spring Harbor, NY 11724, USA
REFERENCE   2  (bases 1 to 179410)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
            clone:OJ148 b05
  JOURNAL   Published only to database (2001) In press
REFERENCE   3  (bases 1 to 179410)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20 AUG 2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (e-mail:rgadb@dna.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81 298 38 7413, Fax:81 298 38 7408)
COMMENT     The draft of the sequence of this BAC clone was generated by
            Monsanto; Monsanto and RGP Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * Note: this is a 'working draft' sequence,
            * this sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..179410
                     /organism="Oryza sativa"
                     /mol_type="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="2"
                     /clone="OJ148 b05"
BASE COUNT    41,54  a  47264  c  47,54  g  41941  t     2  others
ORIGIN
```

```
Percent Similarity: 96.20%  Percent Identity: 96.20%

Matches 1 to 1 : 1

Alignment 1/1  to AC087771  from 1  to 60560

      1/1 GcccgtgtgtgtgtcgtgtnSeqIeqaPtcdcdeteteqtgtcgtgtcSeqIeqAt 54
          ||||| | | |||||||||| ||||||| | | | ||| | |||
    54/54 GCGCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1/1

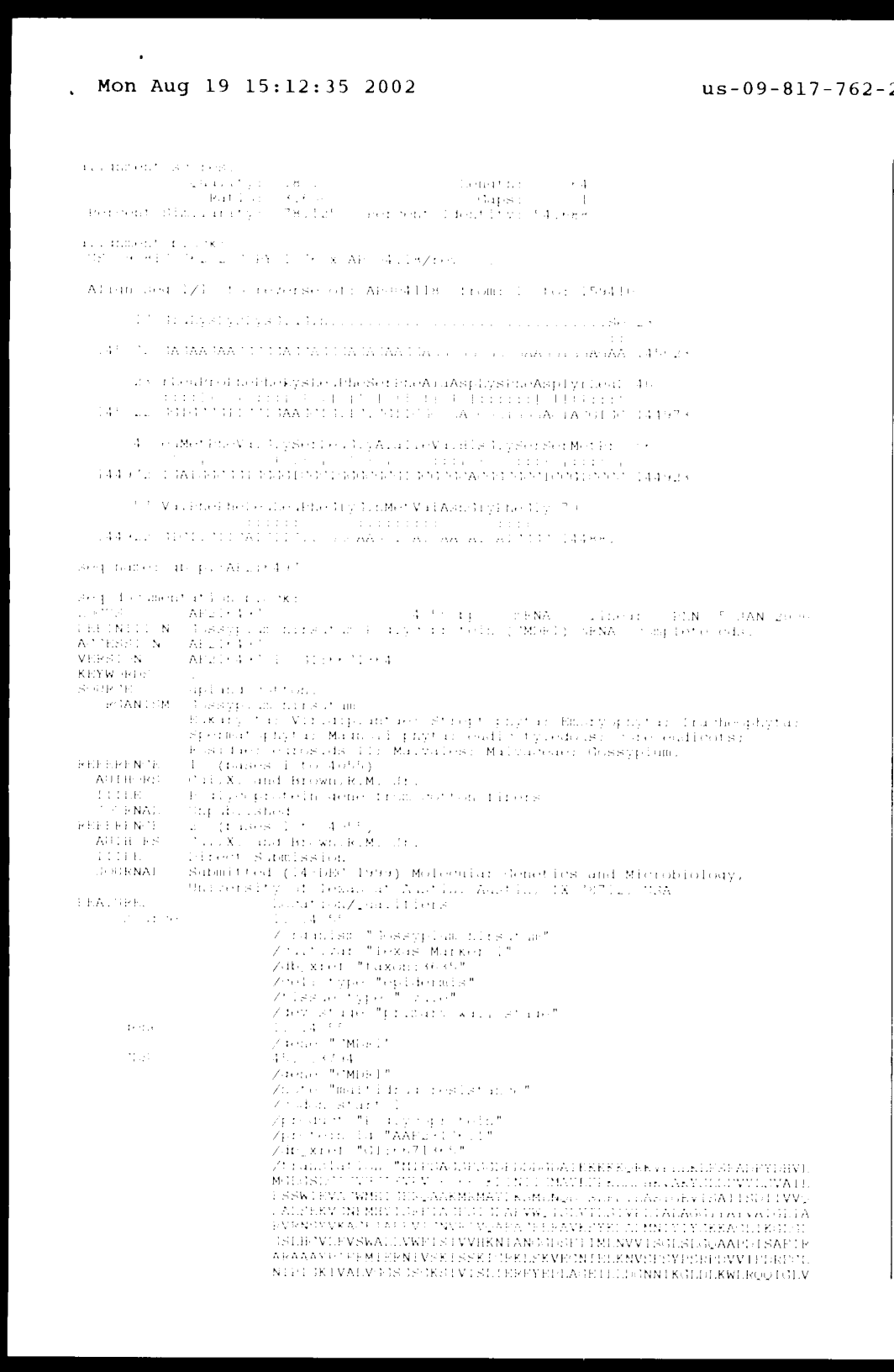
      55 nAspIysIteAspIysIleLeuMetIleaVal?IyaSerLeu?AspLeuV 94
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1/1 NAPYSITEASPLYSILELEUMETILEVAL?LYSSERLEU?ASPLEUV 1/1

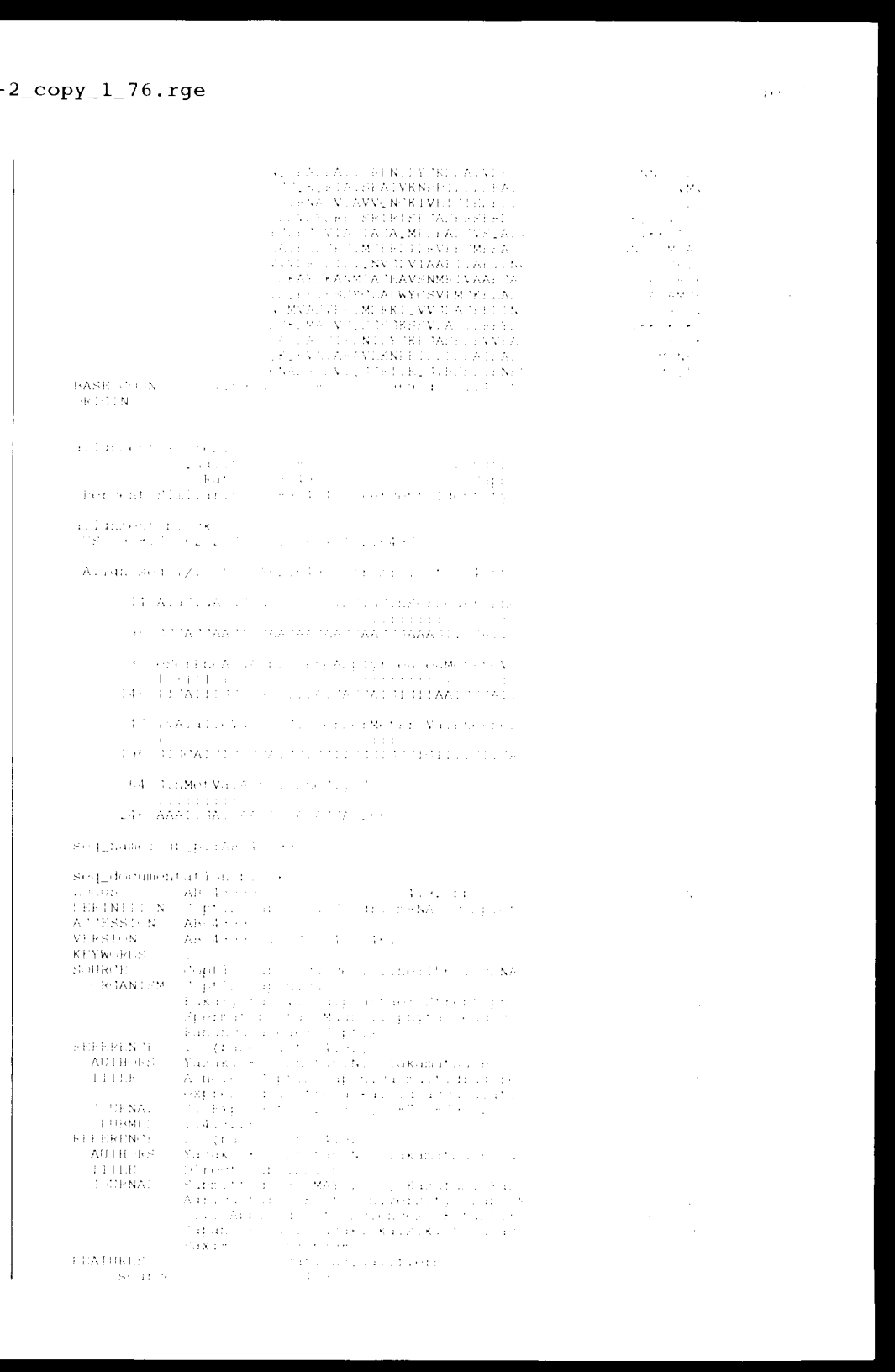
      95 nHisGlySerSerMetProValIheIheIeuPheGlyIhuMetVal 134
         |  | |||||| | | | | | | | | | | | | | | | | | |
    1/1 N?VSLEUASPLYSSERPROVALPHEPHELEUPHEGLYTHRMETVAL 1/1

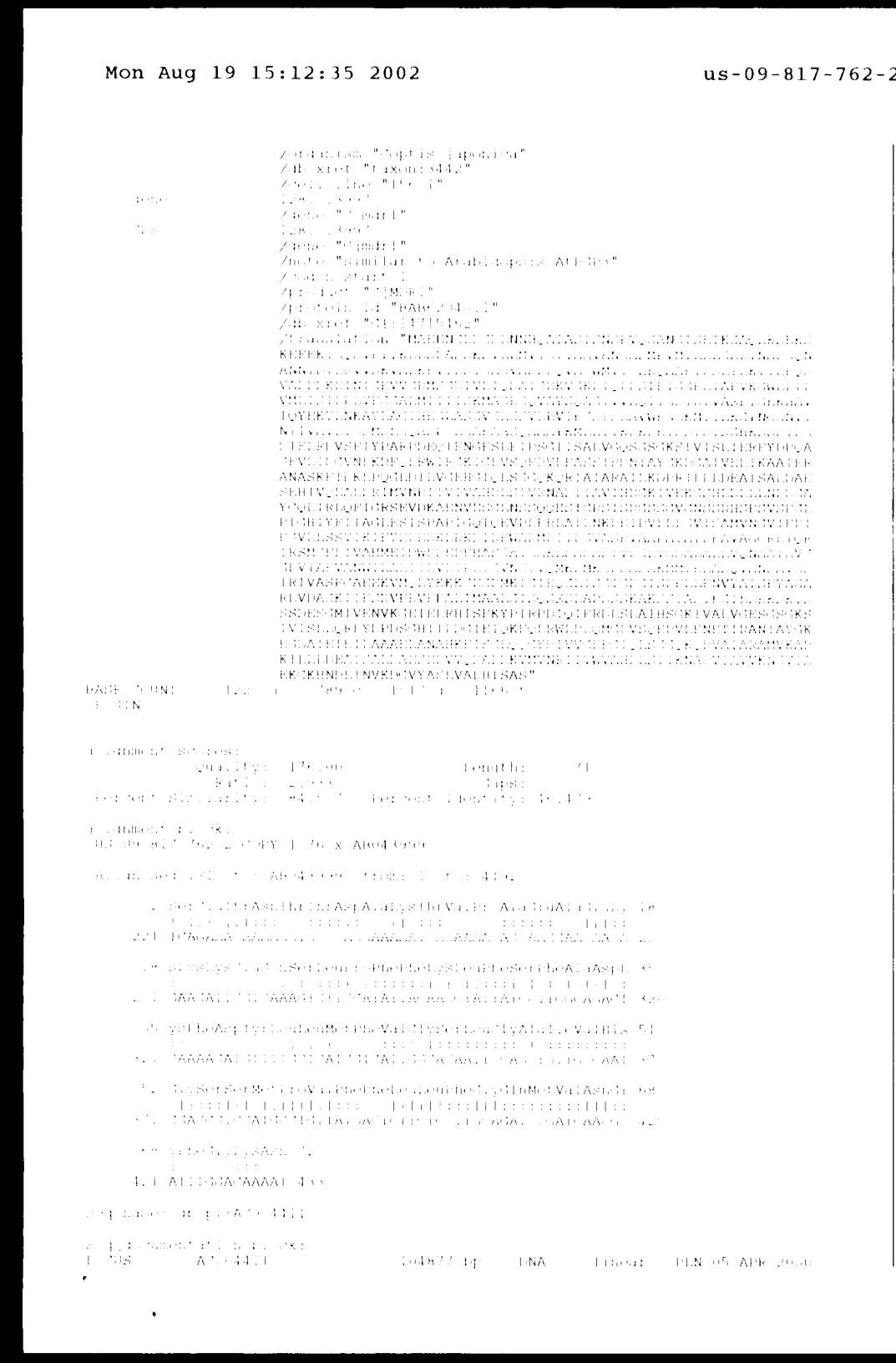
      1/1 AspHyEthoGly 7
          | |  ||| |
    1/1 AATTCEATTTT CTCG

Sequence 1 of 1:AP 41.7

Sequence with gaps:
LOCUS       AP004118               179410 bp    DNA     linear   HTG 7-AUG-2001
DEFINITION  Oryza sativa chromosome 2 clone:OJ148 b05, *** SEQUENCING IN
            PROGRESS ***, 11 ordered pieces.
ACCESSION   AP004118
VERSION     AP004118.1  GI:23475104
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ148 b05.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 179410)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
            clone:OJ148 b05
  JOURNAL   Published only to database (2001) In press
REFERENCE   2  (bases 1 to 179410)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20 AUG 2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (e-mail:rgadb@dna.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81 298 38 7413, Fax:81 298 38 7408)
COMMENT     The draft of the sequence of this BAC clone was generated by
            Monsanto; Monsanto and RGP Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * Note: this is a 'working draft' sequence,
            * this sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..179410
                     /organism="Oryza sativa"
                     /sub_clone="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="2"
                     /clone="OJ148 b05"
BASE COUNT    41,54  a  47264  c  47,54  g  41941  t     2  others
ORIGIN
```







DEFINITION Arabidopsis thaliana chromosome II section 250 of 255 of the complete sequence. Sequence from clones F10C11, F14M4, T5D7
ACCESSION AC004411 AP000554
VERSION AC004411.2 GI:3064
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE 1 (bases 1 to 104877)
AUTHORS Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D., Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J., Ronning,C.M., Koo,H.L., Moffat,K.S., Cronin,L.A., Shen,M., Pai,G., Van Aken,S., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20081567
PUBMED 10617197
REFERENCE 2 (bases 1 to 104877)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REMARK Sequence update by submitter Version replaced gi:4502932. The sequence and annotation of chromosome 2 were derived from those of the individual clones of this chromosome after removal of vector sequences. For additional information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GAIL (ftp://ftp.cbil.upenn.edu/pub/GAIL), GeneMark (Chris Burge, http://genemark.stanford.edu/GeneMark/), and NetPlantGene (http://www.cbs.dtu.dk/services/NetPGene/), searches of the complete sequence against a peptide database and plant EST databases at TAIR, and manual curation based on these analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes either protein or EST similarity, that are predicted by two or more gene prediction programs are most of their models are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Genes were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosomes.

We thank the CSHL/Washu/ABI consortium for sequencing BAC clones F4I2, F5O8, T17A5, and T15I6, the EASA group for sequencing clone F14M4, and Scott Jackson, Jimmie Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics department, especially Lisa Zhou, Hazil Khalak, Michael E. Bowers, Jing Pu, Peter Liang, Jeremy Peterson, Michael Holmes, and Dellwood Richardson for software and database support.

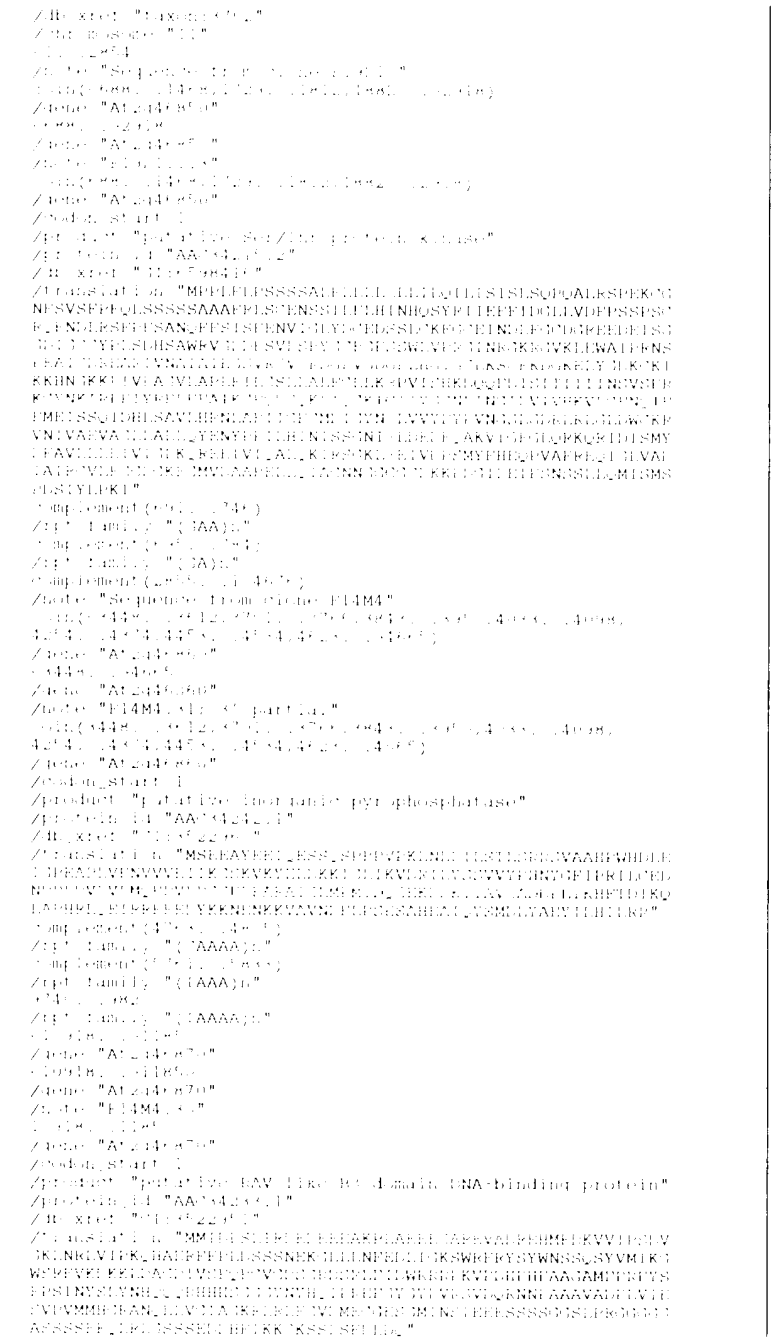
This work was supported by the National Science Foundation, Department of Energy and the USDA Department of Agriculture.

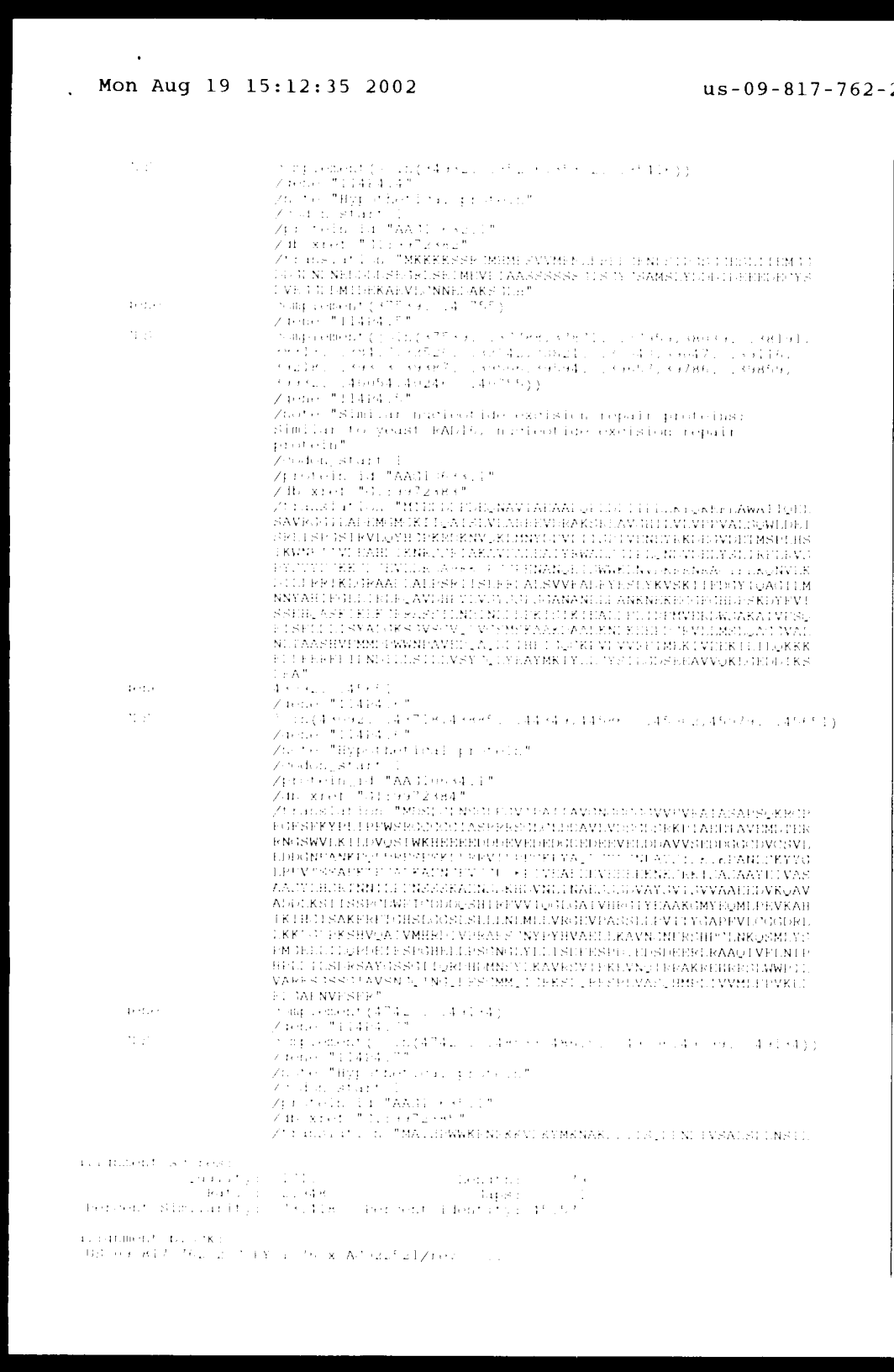
Address all correspondence to Jonathan Lin.
FEATURES Location/Qualifiers
source 1..104877
/organism "Arabidopsis thaliana"
/clone_lib "CSHL/WASHU/ABI"
/clone "F10C11, F14M4, T5D7"
/sex "mixed"
/cultivar "Columbia"
/db_xref "taxon:3702"
/chromosome "II"
/map "II"
gene 1..>1000
/gene "ORF1"
/db_xref "GDB:G1"
CDS 1..>1000
/gene "ORF1"
/codon_start 1
/product "unknown protein"
/protein_id "AAD20000.1"
/db_xref "GI:4321000"
/translation "MSEENIQS
KEPEK
AAG
VGIVKEIGNGIVVVKM
VGLVKEIGNGIVVVKM
IQYEKTGKEAVAVGDDVAFA
NGIVVVKM
GIEGFLVSFFIYPAKDDGTNGESLCISCGLTSALVGQSSGKSVISLIEPYTDDQA
PVLSVQVNLKDPQIPWFKPRDLVSVGEVLEAGSVPENIAYRGGNIVELIKAATEP
ANASKFILKLPQGLDLVVEPHQLSSVGKKTAIARALKGPKPLLILDEAISALDAE
SEHIVQAVLKHMVNGRTVVVVSHDLSVVNAATYLVAIRDGKIVGSGDLDTVKDQV
YVQGLRDQFTGKSEVDKAHNVISSGLNEQQRHTNGTSGPSVVGRGSRFEGVSPD
PDIRIYPFTAGLESISPAFCQGLQEVRLFRLAIGNRTLEPVLLIVIAMVNDVVPT
FSVLLSSSTRTPVSLRLGKCSGWGMCVVVDAVKKGNMKLRVEDPRDLSKSMPIV
IKSMSPSIVAHMKLAWCIGPRNSQAIAAAGIIPLVLDGSSSKDVVVVLAALEVI
RIVASPCAEEEVPLIPEKVDMRLIELLTTMAALVSGGAPSQGRAPCVDAGIMLVD
KLVDAGKIPGDVPEVSEVIAEGTMAALVSGGAPSQGRAPCVDAGIMLVDKLVDAG
SSDESSMIVENVKGEIELPHISFKYPTRPDVLFRDLSLAIHSGKTVALVGESGSGK
IVVSLLQRFYLPDSGHILIDGIDIKDLKLKWLRSQIGVVSQEPVLFNDTIRANIAYG
DSATEIEIIAAAELANAHEFISQLPDGYDTVVGERGVQLSGGQKQRIAIARAIVKN
KILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTVRNADVIAGFDGGV
EKGNHDELMNVKDGVYASLVALHTSAS"
BASE COUNT 1... a ...9... t... c... g... others
ORIGIN

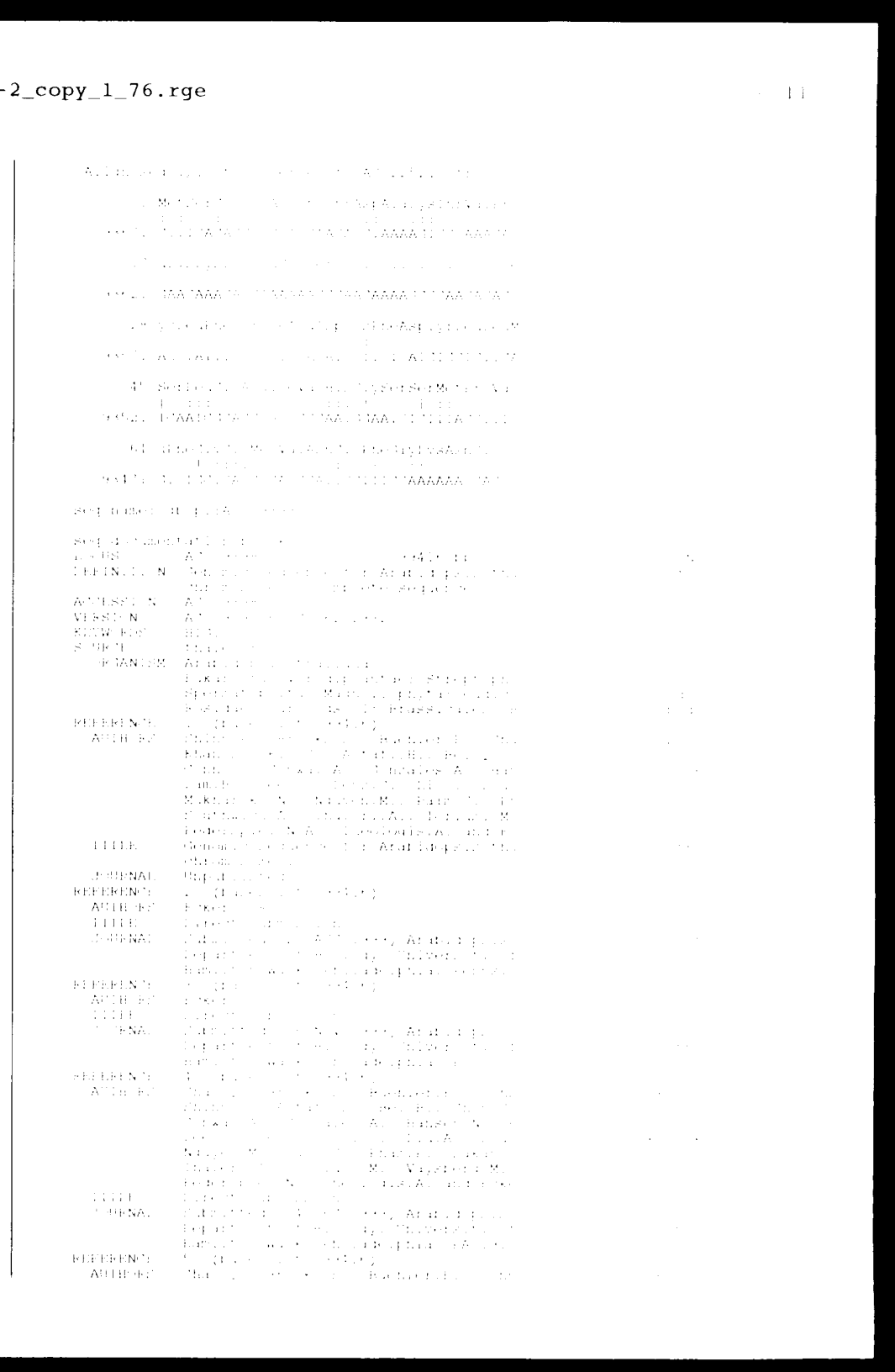
source 1..104877
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="II"
/map="II"
/clone="F10C11, F14M4, T5D7"
/sex="mixed"
/cultivar="Columbia"
gene 1..>1000
/gene="ORF1"
/db_xref="GDB:G1"
CDS 1..>1000
/gene="ORF1"
/codon_start=1
/product="unknown protein"
/protein_id="AAD20000.1"
/db_xref="GI:4321000"
/translation="MSEENIQSKEPEK
VGIVKEIGNGIVVVKM
IQYEKTGKEAVAVGDDVAFA
NGIVVVKM
GIEGFLVSFFIYPAKDDGTNGESLCISCGLTSALVGQSSGKSVISLIEPYTDDQA
PVLSVQVNLKDPQIPWFKPRDLVSVGEVLEAGSVPENIAYRGGNIVELIKAATEP
ANASKFILKLPQGLDLVVEPHQLSSVGKKTAIARALKGPKPLLILDEAISALDAE
SEHIVQAVLKHMVNGRTVVVVSHDLSVVNAATYLVAIRDGKIVGSGDLDTVKDQV
YVQGLRDQFTGKSEVDKAHNVISSGLNEQQRHTNGTSGPSVVGRGSRFEGVSPD
PDIRIYPFTAGLESISPAFCQGLQEVRLFRLAIGNRTLEPVLLIVIAMVNDVVPT
FSVLLSSSTRTPVSLRLGKCSGWGMCVVVDAVKKGNMKLRVEDPRDLSKSMPIV
IKSMSPSIVAHMKLAWCIGPRNSQAIAAAGIIPLVLDGSSSKDVVVVLAALEVI
RIVASPCAEEEVPLIPEKVDMRLIELLTTMAALVSGGAPSQGRAPCVDAGIMLVD
KLVDAGKIPGDVPEVSEVIAEGTMAALVSGGAPSQGRAPCVDAGIMLVDKLVDAG
SSDESSMIVENVKGEIELPHISFKYPTRPDVLFRDLSLAIHSGKTVALVGESGSGK
IVVSLLQRFYLPDSGHILIDGIDIKDLKLKWLRSQIGVVSQEPVLFNDTIRANIAYG
DSATEIEIIAAAELANAHEFISQLPDGYDTVVGERGVQLSGGQKQRIAIARAIVKN
KILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTVRNADVIAGFDGGV
EKGNHDELMNVKDGVYASLVALHTSAS"

BASE COUNT 1... a ...9... t... c... g... others
ORIGIN

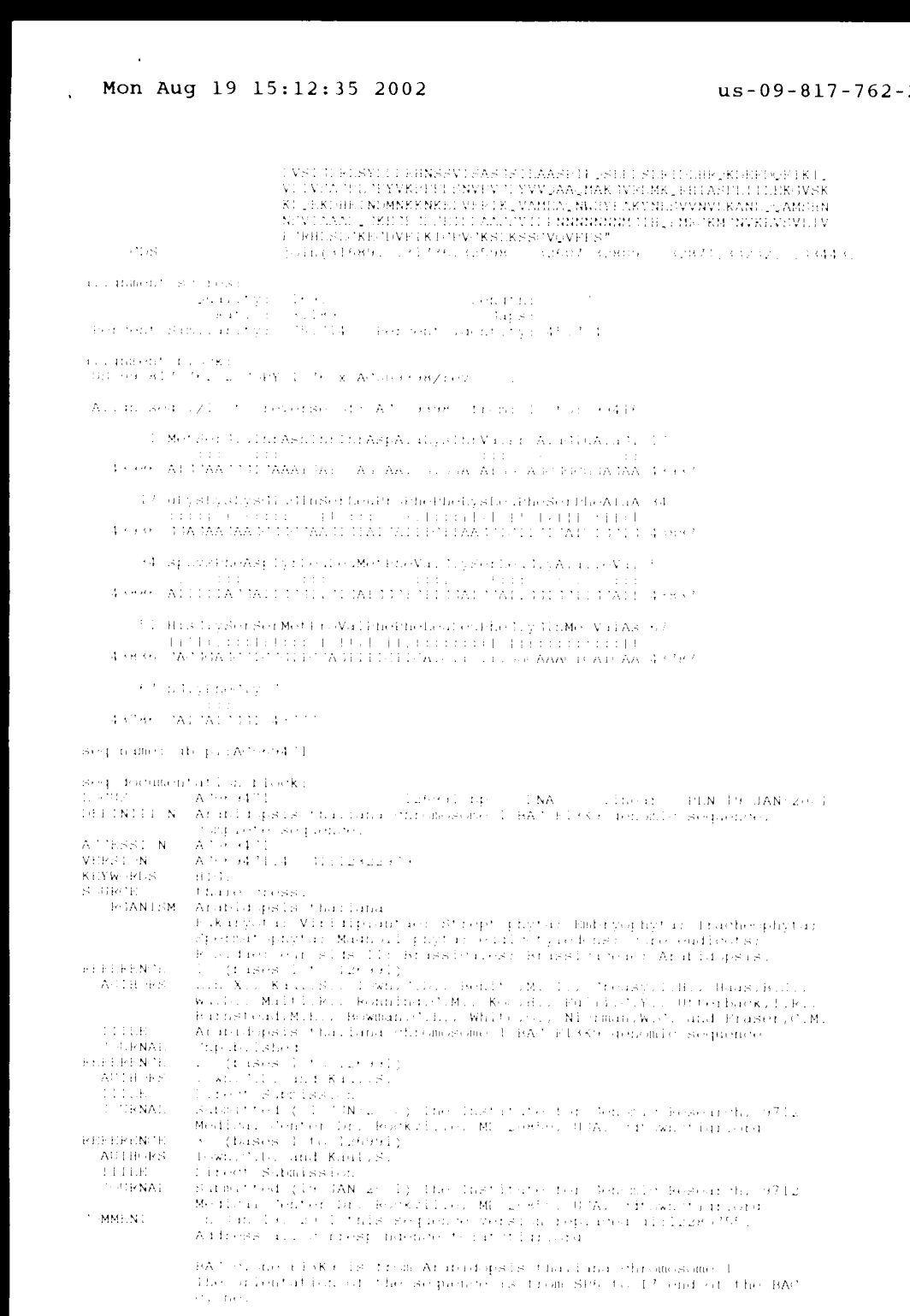
1 atgtcggaag agaacatcca gtcgaaagaa ccggagaaag tcggtatcgt caaagagatc
61 ggaaacggga tcgtcgtcgt caagatggtc ggtctcgtca aagagatcgg aaacgggatc
121 gtcgtcgtca agatggtcgg tctcgtcaaa gagatcggaa acgggatcgt cgtcgtcaag
181 atgatccaat acgagaagac cggaaaagaa gtcgccgtcg gagacgacgt cgccttcgcc
241 aacggaatcg tcgtcgtcaa gatgtgaggt ttcctcgtct ccttcttcat ctacccggcc
301 aaagacgacg gaaccaacgg agagtctctc tgcatctcct gcggtctcac ctccgccctc
361 gtcggccagt cctccggaaa gtccgtcatc tccctcatcg agccctacac cgacgaccag
421 gccccggtcc tctccgtcca agtcaacctc aaagacccgc agatcccgtg gttcaagccg
481 agagacctcg tctccgtcgg agaagtcctc gaagccggat ccgtcccgga gaacatcgcc
541 tacagaggcg gaaacatcgt cgagctcatc aaagccgcca ccgagccggc caacgcctcc
601 aagttcatcc tcaagctccc gcagggactc gacctcgtcg tcgagccgca ccagctctcc
661 tccgtcggaa agaagaccgc catcgccaga gccctcaagg gaccgaagcc gctcctcatc
721 ctcgacgagg ccatctccgc cctcgacgcc gagtccgagc acatcgtcca ggccgtcctc
781 aagcacatgg tcaacggaag aaccgtcgtc gtcgtctccc acgacctctc cgtcgtcaac
841 gccgccacct acctcgtcgc catcagagac ggaaagatcg tcggatccgg agacctcgac
901 accgtcaaag accaggtcta cgtccagggt ctcagagacc agttcaccgg aaagtccgag
961 gtcgacaaag cccacaacgt catctcctcc ggtctcaacg agcagcagag acacaccaac
1021 ggaacctccg gaccgtcggt cgtcggaaga ggatccagat tcgagggagt ctccccggac
1081 ccggacatca gaatctaccc gttcaccgcc ggtctcgagt ccatctcccc ggcgttctgc
1141 caaggtctcc aagaagtcag actcttcaga ctcgccatcg gaaacagaac cctcgagccg
1201 gtcctcctca tcgtcatcgc catggtcaac gacgtcgtcc cgacgttctc cgtcctcctc
1261 tcctcctcca cgagaacgcc ggtctcgctc agactcggaa agtgctccgg atggggaatg
1321 tgcgtcgtcg tcgacgccgt caagaaaggc aacatgaagc tcagagtcga agacccgaga
1381 gacctctcca agtccatgcc gatcgtcatc aagtccatgt ccccgtccat cgtcgcccac
1441 atgaagctcg cgtggtgcat cggaccgaga aactcccaag cgatcgcggc ggcgggaatc
1501 atcccgctcg tcctcgacgg atcctcctcc aaagacgtcg tcgtcgtcct cgcggcgctc
1561 gaagtcatca gaatcgtcgc gtcgccgtgc gcggaagaag aagtcccgct catcccggaa
1621 aaggtcgaca tgagactcat cgagctcctc acgacgatgg cggcgctcgt ctccggcggc
1681 gcgccgtccc agggaagagc gccgtgcgtc gacgcgggaa tcatgctcgt cgacaagctc
1741 gtcgacgccg gaaagatccc gggagacgtc ccggaagtct ccgaagtcat cgcggaaggc
1801 acgatggcgg cgctcgtctc cggcggcgcg ccgtcccagg gaagagcgcc gtgcgtcgac
1861 gcgggaatca tgctcgtcga caagctcgtc gacgccggaa agatcccggg agacgtcccg
1921 gaagtctccg aagtcatcgc ggaaggcacg atggcggcgc tcgtctccgg cggcgcgccg
1981 tcccagggaa gagcgccgtg cgtcgacgcg ggaatcatgc tcgtcgacaa gctcgtcgac
2041 gccggaaaga tcccgggaga cgtcccggaa gtctccgaag tcatcgcgga aggcacgatg
2101 gcggcgctcg tctccggcgg cgcgccgtcc cagggaagag cgccgtgcgt cgacgcggga
2161 atcatgctcg tcgacaagct cgtcgacgcc ggaaagatcc cgggagacgt cccggaagtc
2221 tccgaagtca tcgcggaagg cacgatggcg gcgctcgtct ccggcggcgc gccgtcccag
2281 ggaagagcgc cgtgcgtcga cgcgggaatc atgctcgtcg acaagctcgt cgacgccgga
2341 aagatcccgg gagacgtccc ggaagtctcc gaagtcatcg cggaaggcac gatggcggcg
2401 ctcgtctccg gcggcgcgcc gtcccaggga agagcgccgt gcgtcgacgc gggaatcatg
2461 ctcgtcgaca agctcgtcga cgccggaaag atcccgggag acgtcccgga agtctccgaa
2521 gtcatcgcgg aaggcacgat ggcggcgctc gtctccggcg gcgcgccgtc ccagggaaga
2581 gcgccgtgcg tcgacgcggg aatcatgctc gtcgacaagc tcgtcgacgc cggaaagatc
2641 ccgggagacg tcccggaagt ctccgaagtc atcgcggaag gcacgatggc ggcgctcgtc
2701 tccggcggcg cgccgtccca gggaagagcg ccgtgcgtcg acgcgggaat catgctcgtc
2761 gacaagctcg tcgacgccgg aaagatcccg ggagacgtcc cggaagtctc cgaagtcatc
2821 gcggaaggca cgatggcggc gctcgtctcc ggcggcgcgc cgtcccaggg aagagcgccg
2881 tgcgtcgacg cgggaatcat gctcgtcgac aagctcgtcg acgccggaaa gatcccggga
2941 gacgtcccgg aagtctccga agtcatcgcg gaaggcacga tggcggcgct cgtctccggc
3001 ggcgcgccgt cccagggaag agcgccgtgc gtcgacgcgg gaatcatgct cgtcgacaag
3061 ctcgtcgacg ccggaaagat cccgggagac gtcccggaag tctccgaagt catcgcggaa
3121 ggcacgatgg cggcgctcgt ctccggcggc gcgccgtccc agggaagagc gccgtgcgtc
3181 gacgcgggaa tcatgctcgt cgacaagctc gtcgacgccg gaaagatccc gggagacgtc
3241 ccggaagtct ccgaagtcat cgcggaaggc acgatggcgg cgctcgtctc cggcggcgcg
3301 ccgtcccagg gaagagcgcc gtgcgtcgac gcgggaatca tgctcgtcga caagctcgtc
3361 gacgccggaa agatcccggg agacgtcccg gaagtctccg aagtcatcgc ggaaggcacg
3421 atggcggcgc tcgtctccgg cggcgcgccg tcccagggaa gagcgccgtg cgtcgacgcg
3481 ggaatcatgc tcgtcgacaa gctcgtcgac gccggaaaga tcccgggaga cgtcccggaa
3541 gtctccgaag tcatcgcgga aggcacgatg gcggcgctcg tctccggcgg cgcgccgtcc
3601 cagggaagag cgccgtgcgt cgacgcggga atcatgctcg tcgacaagct cgtcgacgcc
3661 ggaaagatcc cgggagacgt cccggaagtc tccgaagtca tcgcggaagg cacgatggcg
3721 gcgctcgtct ccggcggcgc gccgtcccag ggaagagcgc cgtgcgtcga cgcgggaatc
3781 atgctcgtcg acaagctcgt cgacgccgga aagatcccgg gagacgtccc ggaagtctcc
3841 gaagtcatcg cggaaggcac gatggcggcg ctcgtctccg gcggcgcgcc gtcccaggga
3901 agagcgccgt gcgtcgacgc gggaatcatg ctcgtcgaca agctcgtcga cgccggaaag
3961 atcccgggag acgtcccgga agtctccgaa gtcatcgcgg aaggcacgat ggcggcgctc
4021 gtctccggcg gcgcgccgtc ccagggaaga gcgccgtgcg tcgacgcggg aatcatgctc
4081 gtcgacaagc tcgtcgacgc cggaaagatc ccgggagacg tcccggaagt ctccgaagtc
4141 atcgcggaag gcacgatggc ggcgctcgtc tccggcggcg cgccgtccca gggaagagcg
4201 ccgtgcgtcg acgcgggaat catgctcgtc gacaagctcg tcgacgccgg aaagatcccg
4261 ggagacgtcc cggaagtctc cgaagtcatc gcggaaggca cgatggcggc gctcgtctcc
4321 ggcggcgcgc cgtcccaggg aagagcgccg tgcgtcgacg cgggaatcat gctcgtcgac
4381 aagctcgtcg acgccggaaa gatcccggga gacgtcccgg aagtctccga agtcatcgcg
4441 gaaggcacga tggcggcgct cgtctccggc ggcgcgccgt cccagggaag agcgccgtgc
4501 gtcgacgcgg gaatcatgct cgtcgacaag ctcgtcgacg ccggaaagat cccgggagac
4561 gtcccggaag tctccgaagt catcgcggaa ggcacgatgg cggcgctcgt ctccggcggc
4621 gcgccgtccc agggaagagc gccgtgcgtc gacgcgggaa tcatgctcgt cgacaagctc
4681 gtcgacgccg gaaagatccc gggagacgtc ccggaagtct ccgaagtcat cgcggaaggc
4741 acgatggcgg cgctcgtctc cggcggcgcg ccgtcccagg gaagagcgcc gtgcgtcgac
4801 gcgggaatca tgctcgtcga caagctcgtc gacgccggaa agatcccggg agacgtcccg
4861 gaagtctccg aagtcatcgc ggaaggcacg atggcggcgc tcgtctccgg cggcgcgccg
4921 tcccagggaa gagcgccgtg cgtcgacgcg ggaatcatgc tcgtcgacaa gctcgtcgac
4981 gccggaaaga tcccgggaga cgtcccggaa gtctccgaag tcatcgcgga aggcacgatg
5041 gcggcgctcg tctccggcgg cgcgccgtcc cagggaagag cgccgtgcgt cgacgcggga
5101 atcatgctcg tcgacaagct cgtcgacgcc ggaaagatcc cgggagacgt cccggaagtc
5161 tccgaagtca tcgcggaagg cacgatggcg gcgctcgtct ccggcggcgc gccgtcccag
5221 ggaagagcgc cgtgcgtcga cgcgggaatc atgctcgtcg acaagctcgt cgacgccgga
5281 aagatcccgg gagacgtccc ggaagtctcc gaagtcatcg cggaaggcac gatggcggcg
5341 ctcgtctccg gcggcgcgcc gtcccaggga agagcgccgt gcgtcgacgc gggaatcatg
5401 ctcgtcgaca agctcgtcga cgccggaaag atcccgggag acgtcccgga agtctccgaa
5461 gtcatcgcgg aaggcacgat ggcggcgctc gtctccggcg gcgcgccgtc ccagggaaga
5521 gcgccgtgcg tcgacgcggg aatcatgctc gtcgacaagc tcgtcgacgc cggaaagatc
5581 ccgggagacg tcccggaagt ctccgaagtc atcgcggaag gcacgatggc ggcgctcgtc
5641 tccggcggcg cgccgtccca gggaagagcg ccgtgcgtcg acgcgggaat catgctcgtc
5701 gacaagctcg tcgacgccgg aaagatcccg ggagacgtcc cggaagtctc cgaagtcatc
5761 gcggaaggca cgatggcggc gctcgtctcc ggcggcgcgc cgtcccaggg aagagcgccg
5821 tgcgtcgacg cgggaatcat gctcgtcgac aagctcgtcg acgccggaaa gatcccggga
5881 gacgtcccgg aagtctccga agtcatcgcg gaaggcacga tggcggcgct cgtctccggc
5941 ggcgcgccgt cccagggaag agcgccgtgc gtcgacgcgg gaatcatgct cgtcgacaag
6001 ctcgtcgacg ccggaaagat cccgggagac gtcccggaag tctccgaagt catcgcggaa
6061 ggcacgatgg cggcgctcgt ctccggcggc gcgccgtccc agggaagagc gccgtgcgtc
6121 gacgcgggaa tcatgctcgt cgacaagctc gtcgacgccg gaaagatccc gggagacgtc
6181 ccggaagtct ccgaagtcat cgcggaaggc acgatggcgg cgctcgtctc cggcggcgcg
6241 ccgtcccagg gaagagcgcc gtgcgtcgac gcgggaatca tgctcgtcga caagctcgtc
6301 gacgccggaa agatcccggg agacgtcccg gaagtctccg aagtcatcgc ggaaggcacg
6361 atggcggcgc tcgtctccgg cggcgcgccg tcccagggaa gagcgccgtg cgtcgacgcg
6421 ggaatcatgc tcgtcgacaa gctcgtcgac gccggaaaga tcccgggaga cgtcccggaa
6481 gtctccgaag tcatcgcgga aggcacgatg gcggcgctcg tctccggcgg cgcgccgtcc
6541 cagggaagag cgccgtgcgt cgacgcggga atcatgctcg tcgacaagct cgtcgacgcc
6601 ggaaagatcc cgggagacgt cccggaagtc tccgaagtca tcgcggaagg cacgatggcg
6661 gcgctcgtct ccggcggcgc gccgtcccag ggaagagcgc cgtgcgtcga cgcgggaatc
6721 atgctcgtcg acaagctcgt cgacgccgga aagatcccgg gagacgtccc ggaagtctcc
6781 gaagtcatcg cggaaggcac gatggcggcg ctcgtctccg gcggcgcgcc gtcccaggga
6841 agagcgccgt gcgtcgacgc gggaatcatg ctcgtcgaca agctcgtcga cgccggaaag
6901 atcccgggag acgtcccgga agtctccgaa gtcatcgcgg aaggcacgat ggcggcgctc
6961 gtctccggcg gcgcgccgtc ccagggaaga gcgccgtgcg tcgacgcggg aatcatgctc
7021 gtcgacaagc tcgtcgacgc cggaaagatc ccgggagacg tcccggaagt ctccgaagtc
7081 atcgcggaag gcacgatggc ggcgctcgtc tccggcggcg cgccgtccca gggaagagcg
7141 ccgtgcgtcg acgcgggaat catgctcgtc gacaagctcg tcgacgccgg aaagatcccg
7201 ggagacgtcc cggaagtctc cgaagtcatc gcggaaggca cgatggcggc gctcgtctcc
7261 ggcggcgcgc cgtcccaggg aagagcgccg tgcgtcgacg cgggaatcat gctcgtcgac
7321 aagctcgtcg acgccggaaa gatcccggga gacgtcccgg aagtctccga agtcatcgcg
7381 gaaggcacga tggcggcgct cgtctccggc ggcgcgccgt cccagggaag agcgccgtgc
7441 gtcgacgcgg gaatcatgct cgtcgacaag ctcgtcgacg ccggaaagat cccgggagac
7501 gtcccggaag tctccgaagt catcgcggaa ggcacgatgg cggcgctcgt ctccggcggc
7561 gcgccgtccc agggaagagc gccgtgcgtc gacgcgggaa tcatgctcgt cgacaagctc
7621 gtcgacgccg gaaagatccc gggagacgtc ccggaagtct ccgaagtcat cgcggaaggc
7681 acgatggcgg cgctcgtctc cggcggcgcg ccgtcccagg gaagagcgcc gtgcgtcgac
7741 gcgggaatca tgctcgtcga caagctcgtc gacgccggaa agatcccggg agacgtcccg
7801 gaagtctccg aagtcatcgc ggaaggcacg atggcggcgc tcgtctccgg cggcgcgccg
7861 tcccagggaa gagcgccgtg cgtcgacgcg ggaatcatgc tcgtcgacaa gctcgtcgac
7921 gccggaaaga tcccgggaga cgtcccggaa gtctccgaag tcatcgcgga aggcacgatg
7981 gcggcgctcg tctccggcgg cgcgccgtcc cagggaagag cgccgtgcgt cgacgcggga
8041 atcatgctcg tcgacaagct cgtcgacgcc ggaaagatcc cgggagacgt cccggaagtc
8101 tccgaagtca tcgcggaagg cacgatggcg gcgctcgtct ccggcggcgc gccgtcccag
8161 ggaagagcgc cgtgcgtcga cgcgggaatc atgctcgtcg acaagctcgt cgacgccgga
8221 aagatcccgg gagacgtccc ggaagtctcc gaagtcatcg cggaaggcac gatggcggcg
8281 ctcgtctccg gcggcgcgcc gtcccaggga agagcgccgt gcgtcgacgc gggaatcatg
8341 ctcgtcgaca agctcgtcga cgccggaaag atcccgggag acgtcccgga agtctccgaa
8401 gtcatcgcgg aaggcacgat ggcggcgctc gtctccggcg gcgcgccgtc ccagggaaga
8461 gcgccgtgcg tcgacgcggg aatcatgctc gtcgacaagc tcgtcgacgc cggaaagatc
8521 ccgggagacg tcccggaagt ctccgaagtc atcgcggaag gcacgatggc ggcgctcgtc
8581 tccggcggcg cgccgtccca gggaagagcg ccgtgcgtcg acgcgggaat catgctcgtc
8641 gacaagctcg tcgacgccgg aaagatcccg ggagacgtcc cggaagtctc cgaagtcatc
8701 gcggaaggca cgatggcggc gctcgtctcc ggcggcgcgc cgtcccaggg aagagcgccg
8761 tgcgtcgacg cgggaatcat gctcgtcgac aagctcgtcg acgccggaaa gatcccggga
8821 gacgtcccgg aagtctccga agtcatcgcg gaaggcacga tggcggcgct cgtctccggc
8881 ggcgcgccgt cccagggaag agcgccgtgc gtcgacgcgg gaatcatgct cgtcgacaag
8941 ctcgtcgacg ccggaaagat cccgggagac gtcccggaag tctccgaagt catcgcggaa
9001 ggcacgatgg cggcgctcgt ctccggcggc gcgccgtccc agggaagagc gccgtgcgtc
9061 gacgcgggaa tcatgctcgt cgacaagctc gtcgacgccg gaaagatccc gggagacgtc
9121 ccggaagtct ccgaagtcat cgcggaaggc acgatggcgg cgctcgtctc cggcggcgcg
9181 ccgtcccagg gaagagcgcc gtgcgtcgac gcgggaatca tgctcgtcga caagctcgtc
9241 gacgccggaa agatcccggg agacgtcccg gaagtctccg aagtcatcgc ggaaggcacg
9301 atggcggcgc tcgtctccgg cggcgcgccg tcccagggaa gagcgccgtg cgtcgacgcg
9361 ggaatcatgc tcgtcgacaa gctcgtcgac gccggaaaga tcccgggaga cgtcccggaa
9421 gtctccgaag tcatcgcgga aggcacgatg gcggcgctcg tctccggcgg cgcgccgtcc
9481 cagggaagag cgccgtgcgt cgacgcggga atcatgctcg tcgacaagct cgtcgacgcc
9541 ggaaagatcc cgggagacgt cccggaagtc tccgaagtca tcgcggaagg cacgatggcg
9601 gcgctcgtct ccggcggcgc gccgtcccag ggaagagcgc cgtgcgtcga cgcgggaatc
9661 atgctcgtcg acaagctcgt cgacgccgga aagatcccgg gagacgtccc ggaagtctcc
9721 gaagtcatcg cggaaggcac gatggcggcg ctcgtctccg gcggcgcgcc gtcccaggga
9781 agagcgccgt gcgtcgacgc gggaatcatg ctcgtcgaca agctcgtcga cgccggaaag
9841 atcccgggag acgtcccgga agtctccgaa gtcatcgcgg aaggcacgat ggcggcgctc
9901 gtctccggcg gcgcgccgtc ccagggaaga gcgccgtgcg tcgacgcggg aatcatgctc
9961 gtcgacaagc tcgtcgacgc cggaaagatc ccgggagacg tcccggaagt ctccgaagtc
10021 atcgcggaag gcacgatggc ggcgctcgtc tccggcggcg cgccgtccca gggaagagcg
10081 ccgtgcgtcg acgcgggaat catgctcgtc gacaagctcg tcgacgccgg aaagatcccg
10141 ggagacgtcc cggaagtctc cgaagtcatc gcggaaggca cgatggcggc gctcgtctcc
10201 ggcggcgcgc cgtcccaggg aagagcgccg tgcgtcgacg cgggaatcat gctcgtcgac
10261 aagctcgtcg acgccggaaa gatcccggga gacgtcccgg aagtctccga agtcatcgcg
10321 gaaggcacga tggcggcgct cgtctccggc ggcgcgccgt cccagggaag agcgccgtgc
10381 gtcgacgcgg gaatcatgct cgtcgacaag ctcgtcgacg ccggaaagat cccgggagac
10441 gtcccggaag tctccgaagt catcgcggaa ggcacgatgg cggcgctcgt ctccggcggc
10501 gcgccgtccc agggaagagc gccgtgcgtc gacgcgggaa tcatgctcgt cgacaagctc
10561 gtcgacgccg gaaagatccc gggagacgtc ccggaagtct ccgaagtcat cgcggaaggc
10621 acgatggcgg cgctcgtctc cggcggcgcg ccgtcccagg gaagagcgcc gtgcgtcgac
10681 gcgggaatca tgctcgtcga caagctcgtc gacgccggaa agatcccggg agacgtcccg
10741 gaagtctccg aagtcatcgc ggaaggcacg atggcggcgc tcgtctccgg cggcgcgccg
10801 tcccagggaa gagcgccgtg cgtcgacgcg ggaatcatgc tcgtcgacaa gctcgtcgac
10861 gccggaaaga tcccgggaga cgtcccggaa gtctccgaag tcatcgcgga aggcacgatg
10921 gcggcgctcg tctccggcgg cgcgccgtcc cagggaagag cgccgtgcgt cgacgcggga
10981 atcatgctcg tcgacaagct cgtcgacgcc ggaaagatcc cgggagacgt cccggaagtc
11041 tccgaagtca tcgcggaagg cacgatggcg gcgctcgtct ccggcggcgc gccgtcccag
11101 ggaagagcgc cgtgcgtcga cgcgggaatc atgctcgtcg acaagctcgt cgacgccgga
11161 aagatcccgg gagacgtccc ggaagtctcc gaagtcatcg cggaaggcac gatggcggcg
11221 ctcgtctccg gcggcgcgcc gtcccaggga agagcgccgt gcgtcgacgc gggaatcatg
11281 ctcgtcgaca agctcgtcga cgccggaaag atcccgggag acgtcccgga agtctccgaa
11341 gtcatcgcgg aaggcacgat ggcggcgctc gtctccggcg gcgcgccgtc ccagggaaga
11401 gcgccgtgcg tcgacgcggg aatcatgctc gtcgacaagc tcgtcgacgc cggaaagatc
11461 ccgggagacg tcccggaagt ctccgaagtc atcgcggaag gcacgatggc ggcgctcgtc
11521 tccggcggcg cgccgtccca gggaagagcg ccgtgcgtcg acgcgggaat catgctcgtc
11581 gacaagctcg tcgacgccgg aaagatcccg ggagacgtcc cggaagtctc cgaagtcatc
11641 gcggaaggca cgatggcggc gctcgtctcc ggcggcgcgc cgtcccaggg aagagcgccg
11701 tgcgtcgacg cgggaatcat gctcgtcgac aagctcgtcg acgccggaaa gatcccggga
11761 gacgtcccgg aagtctccga agtcatcgcg gaaggcacga tggcggcgct cgtctccggc
11821 ggcgcgccgt cccagggaag agcgccgtgc gtcgacgcgg gaatcatgct cgtcgacaag
11881 ctcgtcgacg ccggaaagat cccgggagac gtcccggaag tctccgaagt catcgcggaa
11941 ggcacgatgg cggcgctcgt ctccggcggc gcgccgtccc agggaagagc gccgtgcgtc
12001 gacgcgggaa tcatgctcgt cgacaagctc gtcgacgccg gaaagatccc gggagacgtc
12061 ccggaagtct ccgaagtcat cgcggaaggc acgatggcgg cgctcgtctc cggcggcgcg
12121 ccgtcccagg gaagagcgcc gtgcgtcgac gcgggaatca tgctcgtcga caagctcgtc
12181 gacgccggaa agatcccggg agacgtcccg gaagtctccg aagtcatcgc ggaaggcacg
12241 atggcggcgc tcgtctccgg cggcgcgccg tcccagggaa gagcgccgtg cgtcgacgcg
12301 ggaatcatgc tcgtcgacaa gctcgtcgac gccggaaaga tcccgggaga cgtcccggaa
12361 gtctccgaag tcatcgcgga aggcacgatg gcggcgctcg tctccggcgg cgcgccgtcc
12421 cagggaagag cgccgtgcgt cgacgcggga atcatgctcg tcgacaagct cgtcgacgcc
12481 ggaaagatcc cgggagacgt cccggaagtc tccgaagtca tcgcggaagg cacgatggcg
12541 gcgctcgtct ccggcggcgc gccgtcccag ggaagagcgc cgtgcgtcga cgcgggaatc
12601 atgctcgtcg acaagctcgt cgacgccgga aagatcccgg gagacgtccc ggaagtctcc
12661 gaagtcatcg cggaaggcac gatggcggcg ctcgtctccg gcggcgcgcc gtcccaggga
12721 agagcgccgt gcgtcgacgc gggaatcatg ctcgtcgaca agctcgtcga cgccggaaag
12781 atcccgggag acgtcccgga agtctccgaa gtcatcgcgg aaggcacgat ggcggcgctc
12841 gtctccggcg gcgcgccgtc ccagggaaga gcgccgtgcg tcgacgcggg aatcatgctc
12901 gtcgacaagc tcgtcgacgc cggaaagatc ccgggagacg tcccggaagt ctccgaagtc
12961 atcgcggaag gcacgatggc ggcgctcgtc tccggcggcg cgccgtccca gggaagagcg
13021 ccgtgcgtcg acgcgggaat catgctcgtc gacaagctcg tcgacgccgg aaagatcccg
13081 ggagacgtcc cggaagtctc cgaagtcatc gcggaaggca cgatggcggc gctcgtctcc
13141 ggcggcgcgc cgtcccaggg aagagcgccg tgcgtcgacg cgggaatcat gctcgtcgac
13201 aagctcgtcg acgccggaaa gatcccggga gacgtcccgg aagtctccga agtcatcgcg
13261 gaaggcacga tggcggcgct cgtctccggc ggcgcgccgt cccagggaag agcgccgtgc
13321 gtcgacgcgg gaatcatgct cgtcgacaag ctcgtcgacg ccggaaagat cccgggagac
13381 gtcccggaag tctccgaagt catcgcggaa ggcacgatgg cggcgctcgt ctccggcggc
13441 gcgccgtccc agggaagagc gccgtgcgtc gacgcgggaa tcatgctcgt cgacaagctc
13501 gtcgacgccg gaaagatccc gggagacgtc ccggaagtct ccgaagtcat cgcggaaggc
13561 acgatggcgg cgctcgtctc cggcggcgcg ccgtcccagg gaagagcgcc gtgcgtcgac
13621 gcgggaatca tgctcgtcga caagctcgtc gacgccggaa agatcccggg agacgtcccg
13681 gaagtctccg aagtcatcgc ggaaggcacg atggcggcgc tcgtctccgg cggcgcgccg
13741 tcccagggaa gagcgccgtg cgtcgacgcg ggaatcatgc tcgtcgacaa gctcgtcgac
13801 gccggaaaga tcccgggaga cgtcccggaa gtctccgaag tcatcgcgga aggcacgatg
13861 gcggcgctcg tctccggcgg cgcgccgtcc cagggaagag cgccgtgcgt cgacgcggga
13921 atcatgctcg tcgacaagct cgtcgacgcc ggaaagatcc cgggagacgt cccggaagtc
13981 tccgaagtca tcgcggaagg cacgatggcg gcgctcgtct ccggcggcgc gccgtcccag
14041 ggaagagcgc cgtgcgtcga cgcgggaatc atgctcgtcg acaagctcgt cgacgccgga
14101 aagatcccgg gagacgtccc ggaagtctcc gaagtcatcg cggaaggcac gatggcggcg
14161 ctcgtctccg gcggcgcgcc gtcccaggga agagcgccgt gcgtcgacgc gggaatcatg
14221 ctcgtcgaca agctcgtcga cgccggaaag atcccgggag acgtcccgga agtctccgaa
14281 gtcatcgcgg aaggcacgat ggcggcgctc gtctccggcg gcgcgccgtc ccagggaaga
14341 gcgccgtgcg tcgacgcggg aatcatgctc gtcgacaagc tcgtcgacgc cggaaagatc
14401 ccgggagacg tcccggaagt ctccgaagtc atcgcggaag gcacgatggc ggcgctcgtc
14461 tccggcggcg cgccgtccca gggaagagcg ccgtgcgtcg acgcgggaat catgctcgtc
14521 gacaagctcg tcgacgccgg aaagatcccg ggagacgtcc cggaagtctc cgaagtcatc
14581 gcggaaggca cgatggcggc gctcgtctcc ggcggcgcgc cgtcccaggg aagagcgccg
14641 tgcgtcgacg cgggaatcat gctcgtcgac aagctcgtcg acgccggaaa gatcccggga
14701 gacgtcccgg aagtctccga agtcatcgcg gaaggcacga tggcggcgct cgtctccggc
14761 ggcgcgccgt cccagggaag agcgccgtgc gtcgacgcgg gaatcatgct cgtcgaca

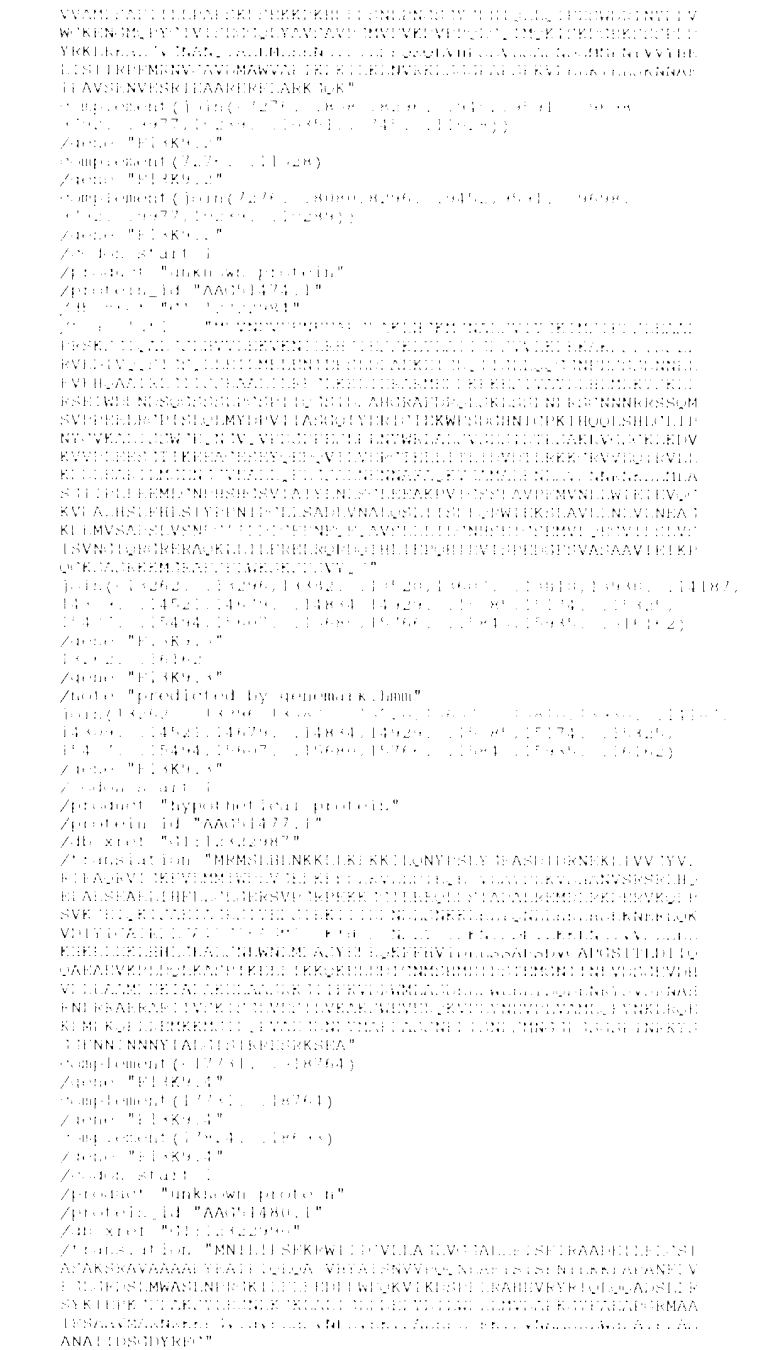


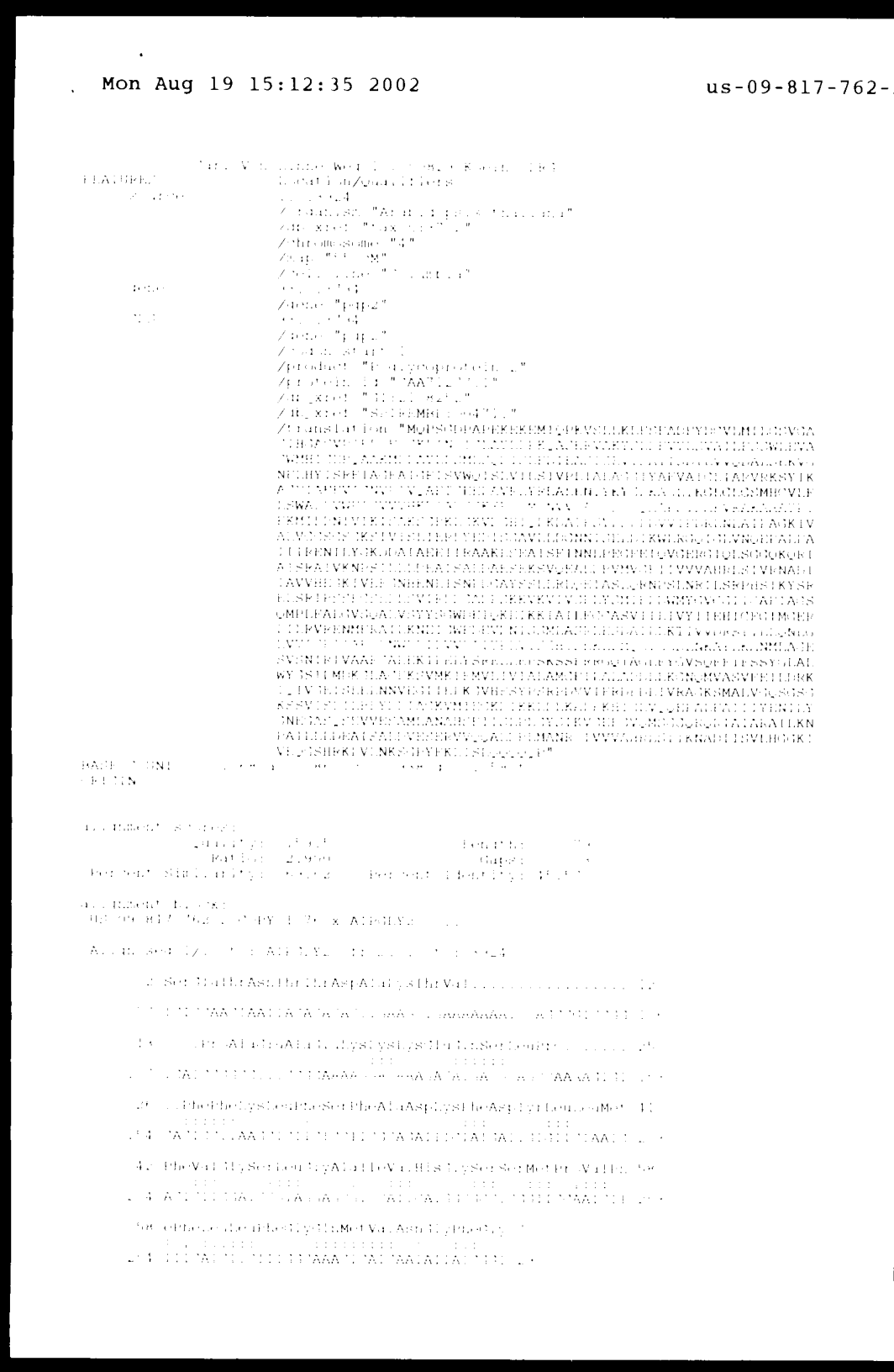


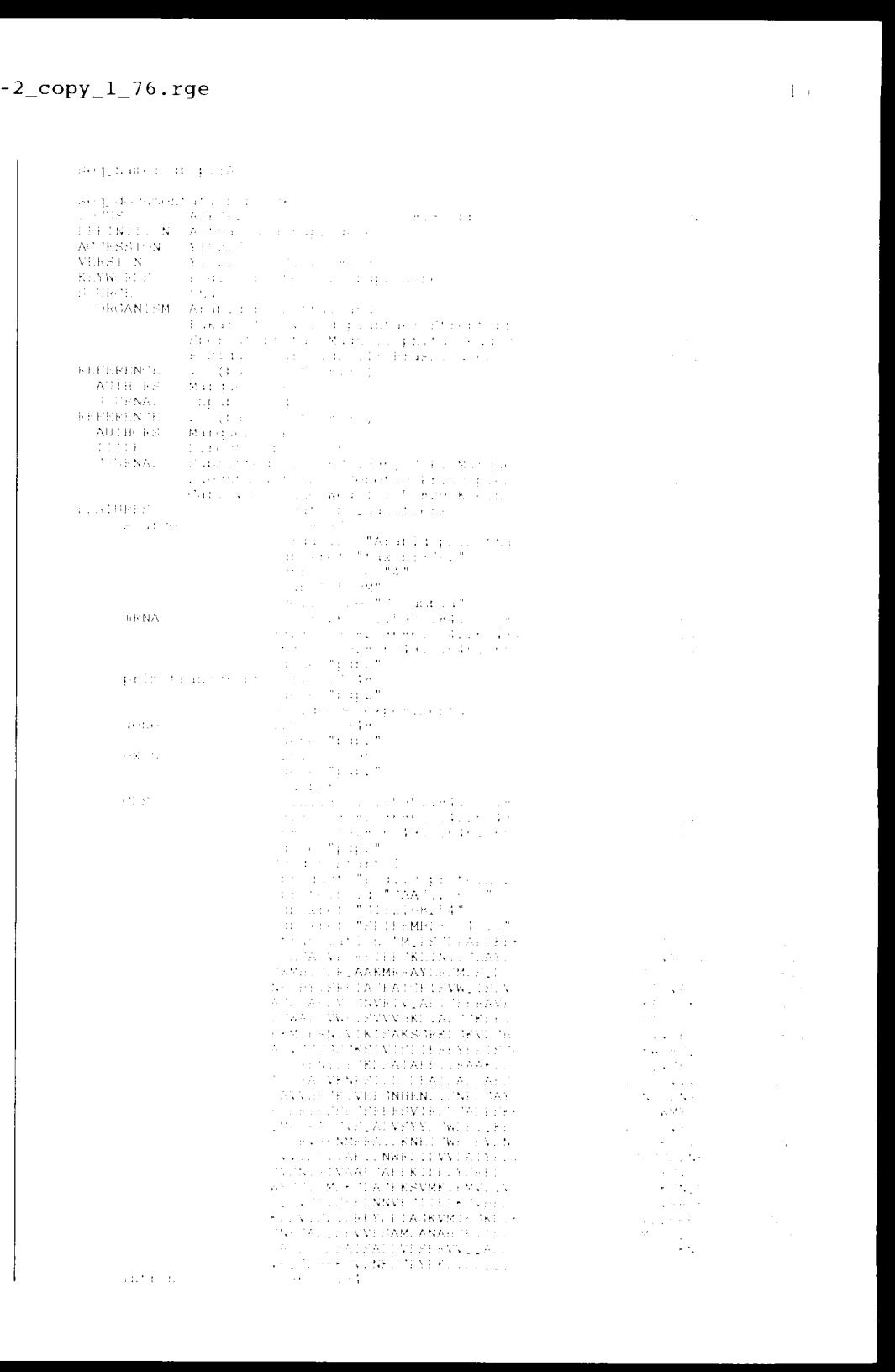


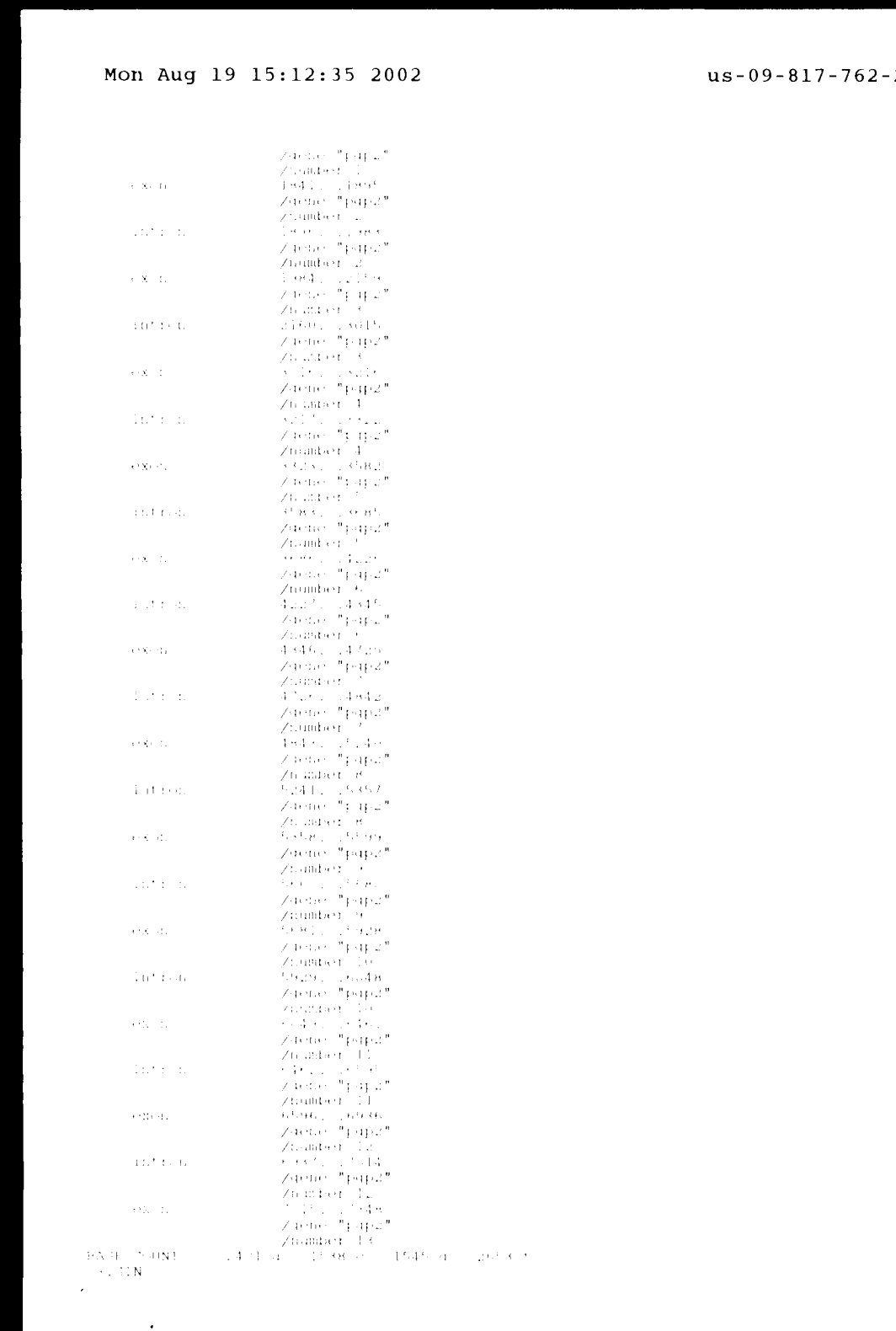
```
            Shinn,P., Altafi,H., Bei,B., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzalez,A., Hansen,N., Howing,B., K[illegible], Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., M[illegible],N.,
            Nguyen,M., Palm,C., Pham,P., [illegible]
            Thaveri,A., Toriumi,M., Vaysberg,M., [illegible], Davis,R.,
            Federspiel,N., Theologis,A. and [illegible]
TITLE       Direct Submission
JOURNAL     Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104 6018, USA
COMMENT     On Nov 6, 1999 this sequence version replaced gi:6137068.
FEATURES             Location/Qualifiers
     source          1..99416
                     /organism "Arabidopsis thaliana"
                     /db_xref "taxon:3702"
                     /chromosome "1"
                     /clone "F20B24"
     CDS             complement(join(2528..2776,2941..3237))
                     /note "similar to tyrosyl tRNA synthetase homolog
                     gi|1707012; similar to ESTs gb|T43206, gb|AI996335.1,
                     gb|H76501, gb|AA605126, gb|N65365"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.1"
                     /protein_id "AAF17[illegible].1"
                     /db_xref "GI:[illegible]"
                     /translation "MEI[illegible]
                     [illegible]
                     [illegible]
                     IIVKDNNNLSLIIYDLVGVEI"
     CDS             join(3969..4197,4232..4317,4372..4596,4638..4692,
                     4773..5040,5099..5394)
                     /note "similar to AMSH gi|4098124"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.2"
                     /protein_id "AAF17[illegible].1"
                     /db_xref "GI:[illegible]"
                     /translation "MV[illegible]
                     ILRDVHIVRTLRPTTE[illegible]
                     AFLN[illegible]
                     SWLP[illegible]
                     LYPVGWIHVYLS[illegible]
                     YQVH[illegible]
                     IFK[illegible]LL"
     CDS             6127..6402
                     /note "similar to C. elegans AL2 homolog gi|1465851"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.3"
                     /protein_id "AAF1765[illegible].1"
                     /db_xref "GI:[illegible]"
                     /translation "M[illegible]
                     LKIVKG[illegible]"
     CDS             join(6667..6901,7022..7194)
                     /note "[illegible]"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.4"
                     /protein_id "AAF17[illegible].1"
                     /db_xref "GI:[illegible]"
                     /translation "MMMM[illegible]
                     SLLFVRCSCESSSAVYYEV[illegible]
                     VFDFNCLWFVLLYHLPPV[illegible]"
     CDS             join(7417..7551,7803..8282,8305..8808)
                     /note "similar to plasmodium [illegible] binding protein [illegible]
                     gi|1142619"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.5"
                     /protein_id "AAF17[illegible].1"
                     /db_xref "GI:[illegible]"
                     /translation "M[illegible]
                     VRPTD[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]VNFTKI[illegible]VNVKVVT"
     CDS             complement(join(9205..9447,9545..9706,9803..9956,
                     10233..10308,10435..10521,10620..11123,11252..12179))
                     /note "putative serine/threonine protein kinase
                     emb|CAA16823.1"
                     /codon_start 1
                     /evidence not experimental
                     /product "F20B24.6"
                     /protein id "AAF17672.1"
                     /db_xref "GI:6573752"
                     /translation "MDKV[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]NEIRIFRRASEDSSD
                     [illegible]NKH"
     CDS             complement(join(13504..13636,13725..13847,13934..14[illegible],
                     14139..14209,14287..14434))
                     /note "similar to ADP ribosylation factor gb|AAD17207;
                     similar to ESTs gb|N65779, gb|AI996946.1, gb|AI725909.1,
                     gb|145928, gb|T42202"
                     /codon_start 1
                     /evidence not experimental
                     /product "F20B24.7"
                     /protein_id "AAF17671.1"
                     /db_xref "GI:6573751"
                     /translation "MGL[illegible]
                     [illegible]
                     [illegible]
                     [illegible]EVKLLNS"
     CDS             complement(join(15682..16265,16344..16551,16591..16665,
                     [illegible],17604..18011))
                     /note "similar to polygalacturonase like protein
                     emb|CAA66811; similar to ESTs gb|R29820, gb|AI993509.1,
                     emb|F19830"
                     /codon_start 1
                     /evidence not_experimental
                     /product "F20B24.8"
                     /protein id "AAF17670.1"
                     /db_xref "GI:6573750"
                     /translation "M[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]VC"
     CDS             complement(join(24[illegible]..25323,25414..25522))
                     /note "similar to putative inhibitor of apoptosis
                     gi|3924605, similar to ESTs gb|AI994578.1, gb|104171,
                     gb|AA728525"
                     /codon start 1
                     /evidence not experimental
                     /product "F20B24.9"
                     /protein_id "AAF17669.1"
                     /db_xref "GI:6573749"
                     /translation "M[illegible]
                     [illegible]
```

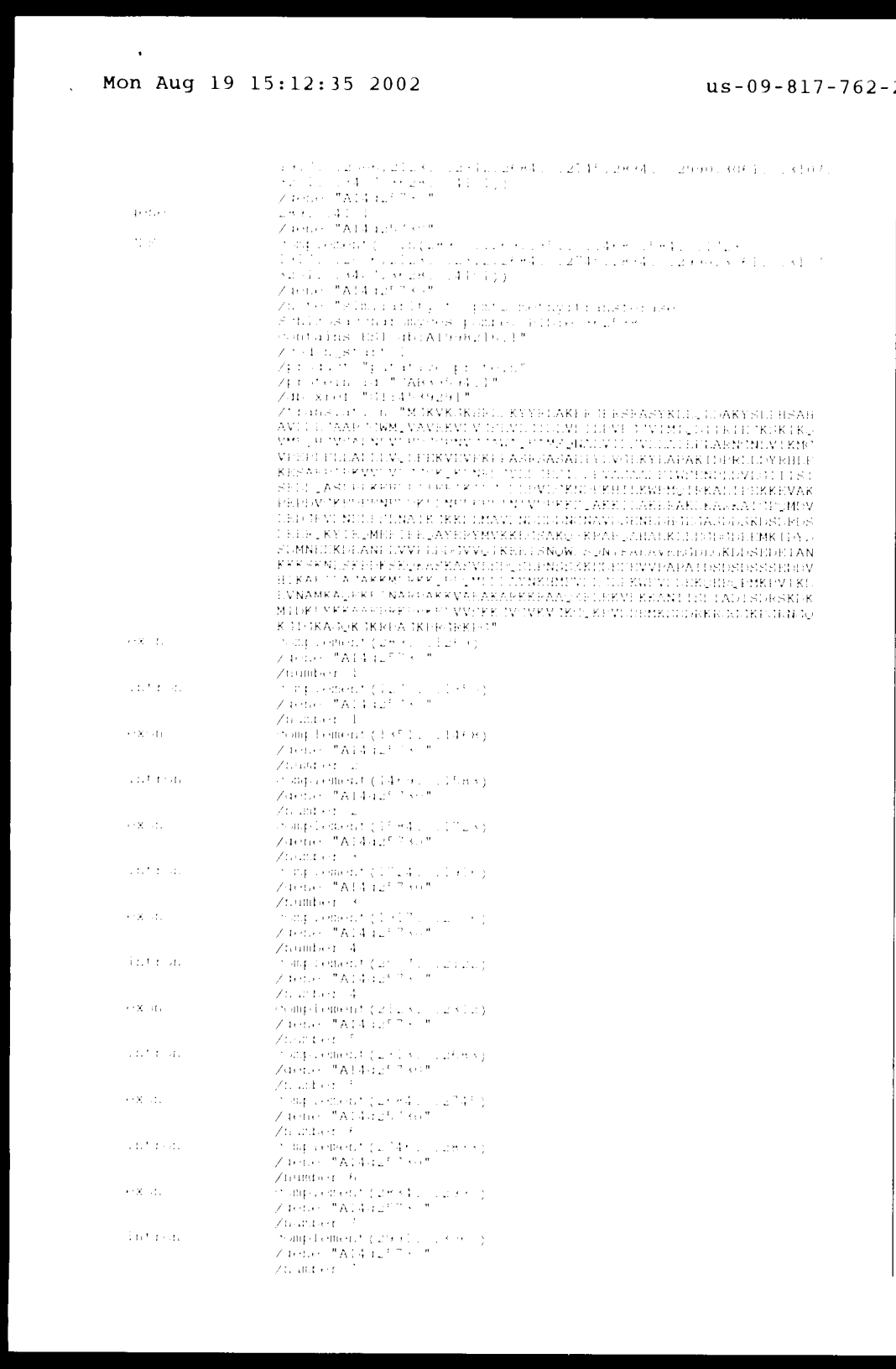


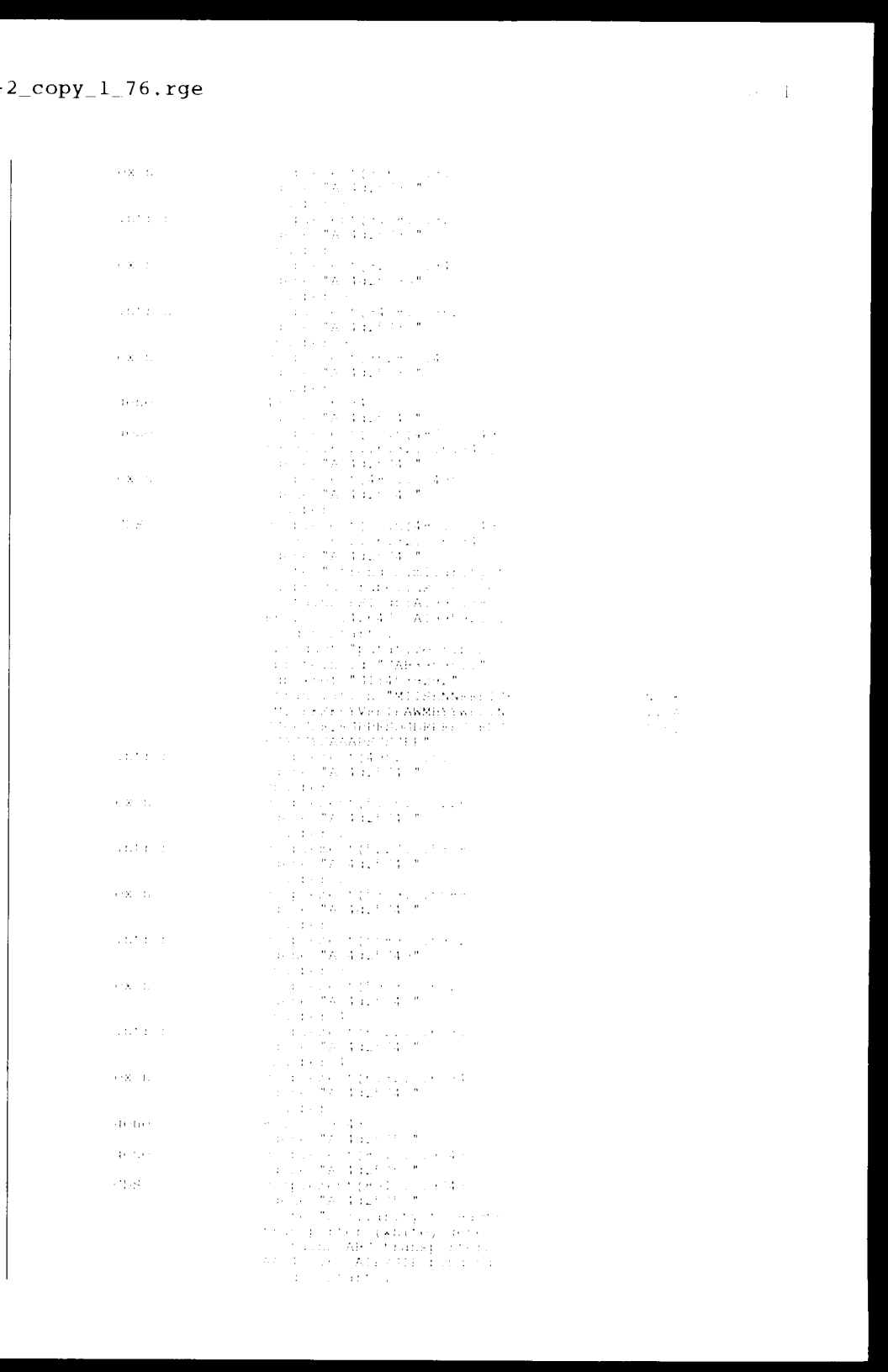


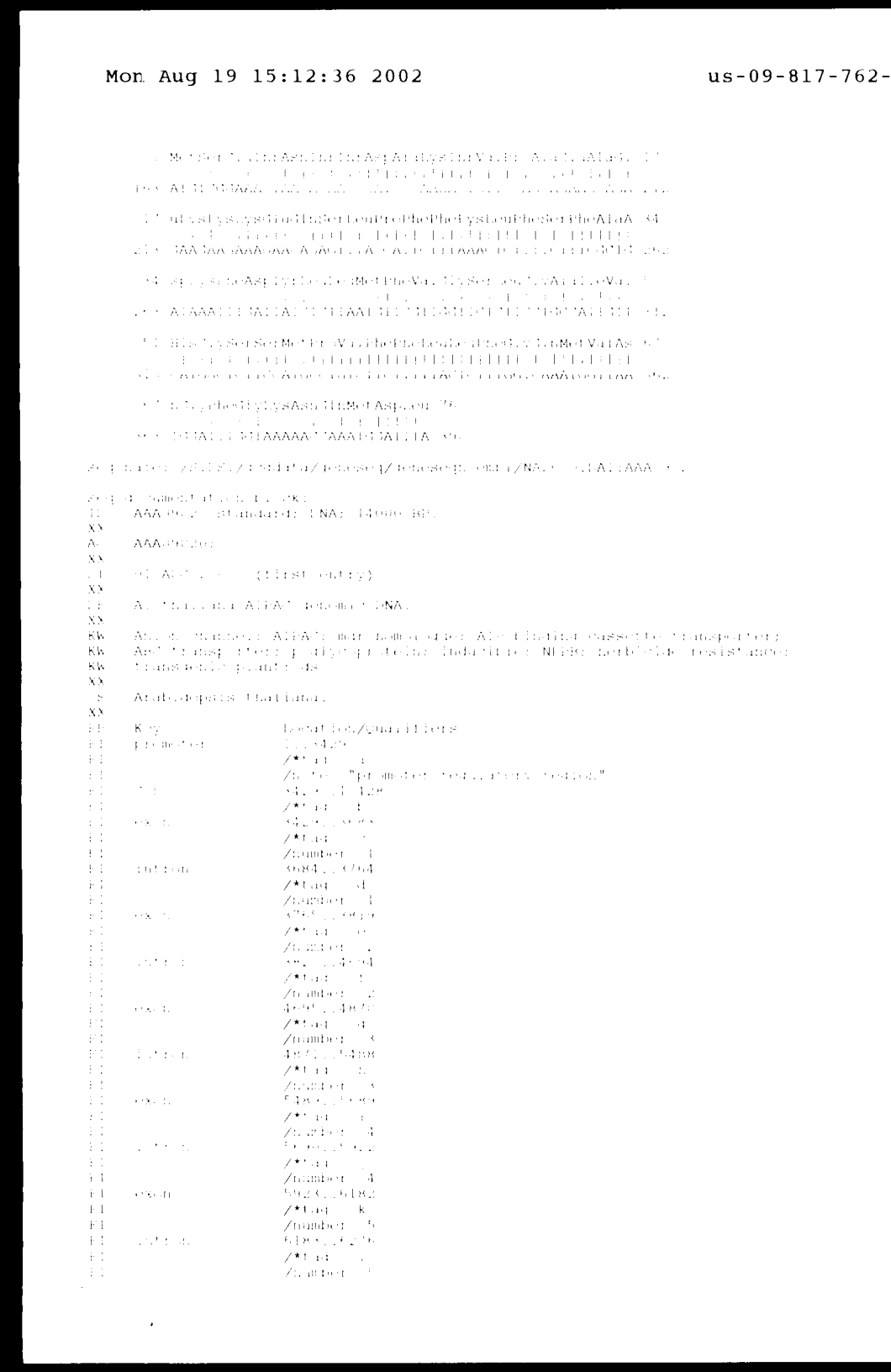


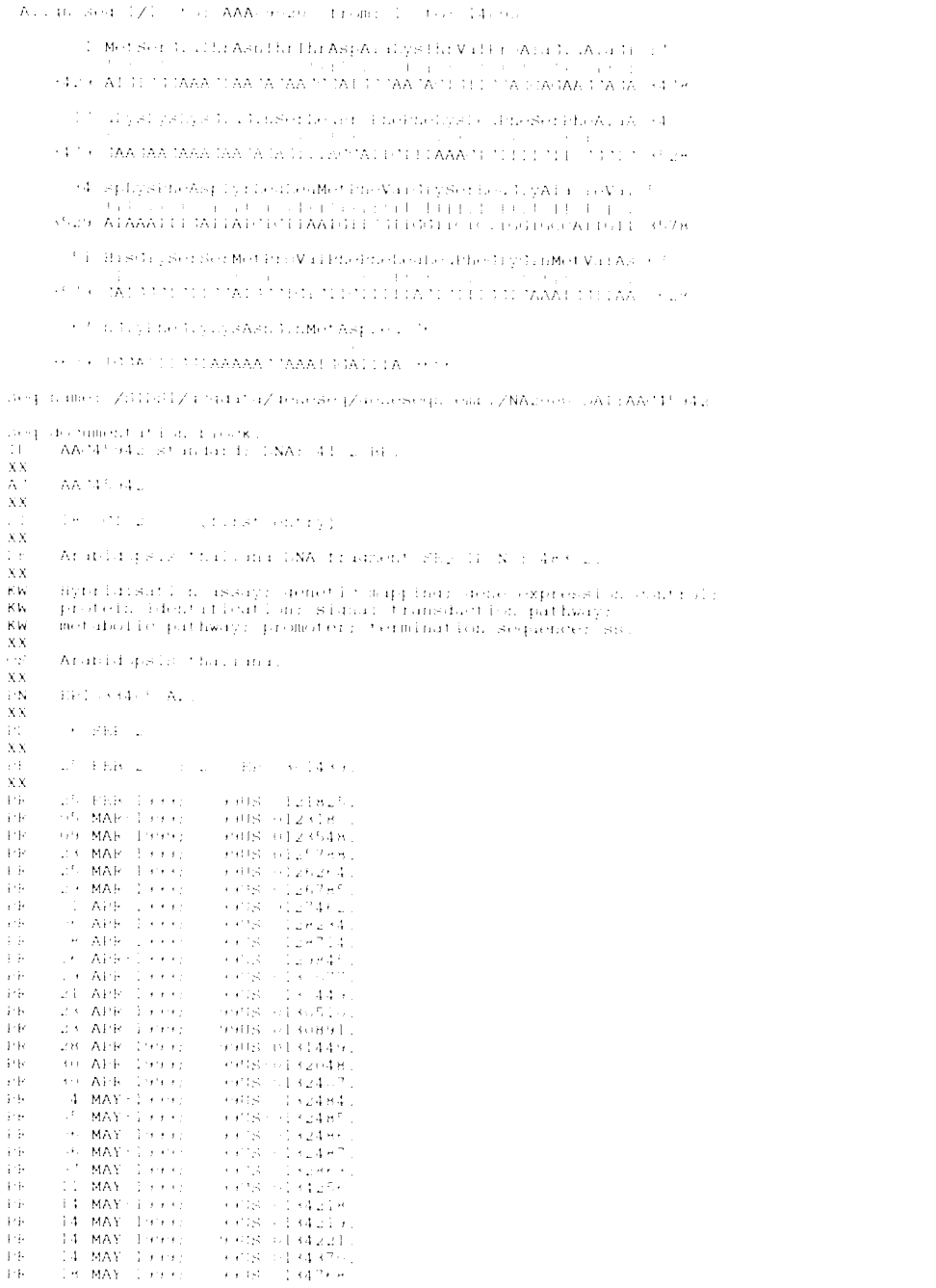












PR 28 JUL 1999; 99US-0145951.
PR 02 AUG 1999; 99US-0146386.
PR 02 AUG 1999; 99US-0146388.
PR 02 AUG 1999; 99US-0146389.
PR 03 AUG 1999; 99US-0147038.
PR 04 AUG 1999; 99US-0147204.
PR 04 AUG 1999; 99US-0147302.
PR 05 AUG 1999; 99US-0147192.
PR 06 AUG 1999; 99US-0147260.
PR 06 AUG 1999; 99US-0147303.
PR 06 AUG 1999; 99US-0147416.
PR 09 AUG 1999; 99US-0147493.
PR 09 AUG 1999; 99US-0147735.
PR 10 AUG 1999; 99US-0148171.
PR 11 AUG 1999; 99US-0148519.
PR 12 AUG 1999; 99US-0148741.
PR 13 AUG 1999; 99US-0148565.
PR 13 AUG 1999; 99US-0148684.
PR 16 AUG 1999; 99US-0149368.
PR 17 AUG 1999; 99US-0149475.
PR 18 AUG 1999; 99US-0149426.
PR 20 AUG 1999; 99US-0149722.
PR 20 AUG 1999; 99US-0149723.
PR 20 AUG 1999; 99US-0149729.
PR 23 AUG 1999; 99US-0149902.
PR 25 AUG 1999; 99US-0149930.
PR 25 AUG 1999; 99US-0150066.
PR 27 AUG 1999; 99US-0150884.
PR 27 AUG 1999; 99US-0151065.
PR 27 AUG 1999; 99US-0151066.
PR 27 AUG 1999; 99US-0151080.
PR 30 AUG 1999; 99US-0151504.
PR 31 AUG 1999; 99US-0151438.
PR 01 SEP 1999; 99US-0151933.
PR 07 SEP 1999; 99US-0152463.
PR 10 SEP 1999; 99US-0154070.
PR 13 SEP 1999; 99US-0154258.
PR 15 SEP 1999; 99US-0154018.
PR 15 SEP 1999; 99US-0154030.
PR 20 SEP 1999; 99US-0154779.
PR 22 SEP 1999; 99US-0155139.
PR 23 SEP 1999; 99US-0155486.
PR 24 SEP 1999; 99US-0155659.
PR 28 SEP 1999; 99US-0156458.
PR 29 SEP 1999; 99US-0156596.
PR 01 OCT 1999; 99US-0157117.
PR 05 OCT 1999; 99US-0157753.
PR 05 OCT 1999; 99US-0157865.
PR 07 OCT 1999; 99US-0158029.
PR 08 OCT 1999; 99US-0158232.
PR 12 OCT 1999; 99US-0158869.
PR 13 OCT 1999; 99US-0159293.
PR 13 OCT 1999; 99US-0159294.
PR 13 OCT 1999; 99US-0159295.
PR 14 OCT 1999; 99US-0159329.
PR 14 OCT 1999; 99US-0159330.
PR 14 OCT 1999; 99US-0159331.
PR 14 OCT 1999; 99US-0159637.
PR 14 OCT 1999; 99US-0159638.
PR 18 OCT 1999; 99US-0159684.
PR 21 OCT 1999; 99US-0160741.
PR 21 OCT 1999; 99US-0160767.
PR 21 OCT 1999; 99US-0160768.
PR 21 OCT 1999; 99US-0160770.
PR 21 OCT 1999; 99US-0160814.
PR 21 OCT 1999; 99US-0160815.
PR 22 OCT 1999; 99US-0160980.
PR 22 OCT 1999; 99US-0160981.
PR 22 OCT 1999; 99US-0160989.
PR 25 OCT 1999; 99US-0161404.
PR 25 OCT 1999; 99US-0161405.
PR 25 OCT 1999; 99US-0161406.
PR 26 OCT 1999; 99US-0161359.

PR 26 OCT 1999; 99US-0161360.
PR 26 OCT 1999; 99US-0161361.
PR 28 OCT 1999; 99US-0161920.
PR 28 OCT 1999; 99US-0161992.
PR 28 OCT 1999; 99US-0161994.
PR 29 OCT 1999; 99US-0162142.

alignment scores:
Quality: 154.50 Length: 63
Ratio: 3.010 Gaps: 1
Percent Similarity: 80.952 Percent Identity: 49.206

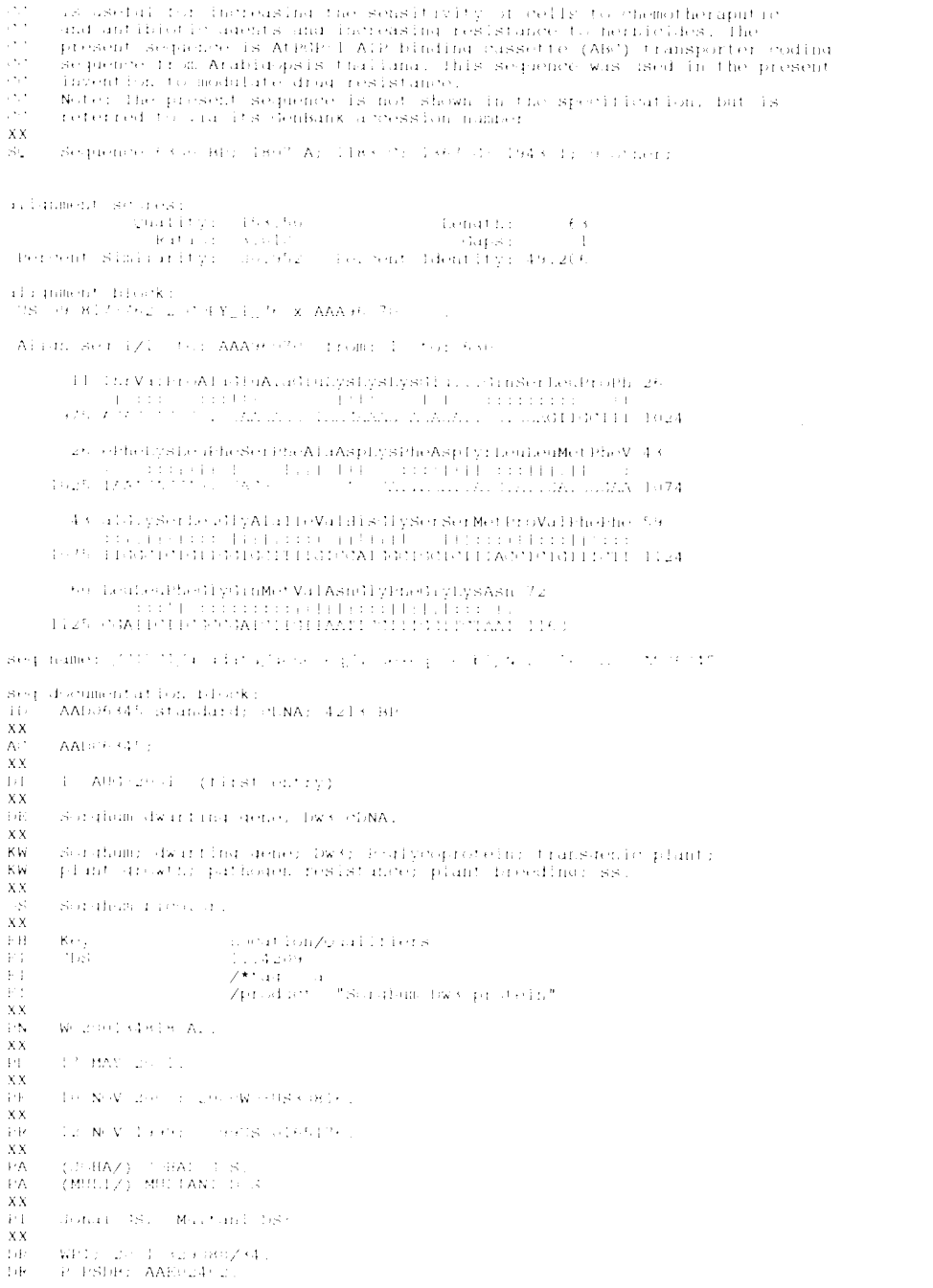
alignment block:
US-09-817-762-2_copy_1_76 x AAC45942 ..

Align seq 171 to AAC45942 from 1 to 4102

11 ThrValProAlaGlnAlaGluLysLysLysGlu...GlnSerLeuProPh 26
26 cPheLysLeuPheSerPheAlaAspLysPheAspTyrLeuLeuMetPheV 43
43 aroGlySerLeuGlyAlaIleValHisGlySerSerMetProValPhePhe 59
60 LeuLeuPheGlyGluMetValAsnGlyPheGlyLysAsn 72

seq_name: /...

seq_documentation block:
ID AAA96070 standard; DNA; 6500 BP.
XX
AC AAA96070;
XX
DT 29 JAN 2001 (first entry)
XX
DE A1PGP1 ATP binding cassette (ABC) transporter coding sequence
XX
KW drug resistance; ATP gradient; chemotherapeutic, antibiotic, herbicide;
KW AtPGP 1; ATP binding cassette; ABC transporter; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200052144 A1.
XX
PD 08 SEP 2000.
XX
PF 28 FEB 2000; 2000WO-US05415.
XX
PR 04 MAR 1999; 99US-0261825.
XX
PA (ISSC) UNIV ILLINOIS.
XX
PI Thomas CL, Winkler JB, Roux SJ, Lloyd AM, Hurley L;
XX
DR WPI; 2000-587306/55.
XX
PT Increasing or decreasing drug resistance in target bacteria, yeast,
PT plant or mammalian cells comprises altering ATP gradient across
PT biological membrane of target cell
XX
PS Claim 15; Page ; 85pp; English.
XX
CC The present invention relates to a method for increasing or decreasing
CC drug resistance in target bacteria, yeast, plant or mammalian cells by
CC altering the ATP gradient across the biological membrane of the target
CC cell. The method is useful for modulating drug resistance of cells. It



```
FT                   /*tag    a
FT                   /number   1
FT   intron          639..885
FT                   /*tag    b
FT                   /number   1
FT   exon            886..1440
FT                   /*tag    c
FT                   /number   2
FT   intron          1441..1950
FT                   /*tag    d
FT                   /number   2
FT   exon            1951..1876
FT                   /*tag    e
FT                   /number   3
FT   intron          1877..2722
FT                   /*tag    f
FT                   /number   3
FT                   /cons_splice  (5'site: YES, 3'site:NO)
FT   exon            2723..2952
FT                   /*tag    g
FT                   /number   4
FT   intron          2953..4423
FT                   /*tag    h
FT                   /number   4
FT   exon            4424..4827
FT                   /*tag    i
FT                   /number   5
XX
PN   WO2001043819-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30821.
XX
PR   12-NOV-1999;  99US-0164886.
XX
PA   (BOHA/) BOHALO S.
PA   (MOLE/) MOLIANI D.S.
XX
PI   Bohal GS,  Moliani DS;
XX
DR   WPI; 2001-326088/34.
XX
PT   Novel sorghum dwarfing gene useful for modifying the growth of the
PT   organisms, particularly plants, such as maize, wheat, rice, Basmati
PT   rice, rye, millet and barley and for enhancing resistance to pathogens
PT   -
XX
PS   Claim 1; Page 58-66; 76pp; English.
XX
CC   The present sequence is full length sorghum dwarfing gene, Dw3.
CC   Dw3 DNA is useful for modifying the growth of an organism, especially
CC   reducing the height of monocotyledon plants, such as maize, wheat, rice,
CC   Basmati rice, sorghum, rye, millet and barley. Dw3 DNA is useful for
CC   producing transgenic monocotyledon and dicotyledon plants like soybean,
CC   sunflower, safflower, alfalfa, cotton, Brassica spp., potatoes or fruit
CC   trees. The Dw3 DNA is also useful for isolating nucleotide homologue
CC   molecules that encode P-glycoproteins, enhancing the resistance of plants
CC   to pathogens, including bacteria, fungi, viruses, nematodes and insects,
CC   as probes for the isolation of other P-glycoprotein like genes and as
CC   molecular markers. Identifying plants that possess a mutant allele is
CC   useful in agriculture, particularly in breeding dwarf crop plants,
CC   particularly dwarf sorghum plants.
XX
SQ   Sequence 4827 BP; 1128 A; 1112 C; 1076 G; 1447 T; 0 other;


Alignment scores:
            Quality:  134.00              Length:     70
              Ratio:  2.627                 Gaps:      0
 Percent Similarity:  72.857   Percent Identity: 41.429

Alignment block:
```

```
US-09-817-762-2_COPY_1_76 x AAW60544     .

Aligned seq 1/1  to: AAW60544  from: 1  to: 5627

       5 IleThrAspAlaLysIleValProAlaGlnAlaGlyLysLysSerH 22
         ||||||||||||  ||||||   |||||   || ||||||
     444 ATCACCGACGCGAAGATCGTCCCGGCGCAGGCGGGCAAGAAGTCCA 488

      22 nSerLeuProPheLysLeuProSerProAlaAspLysIleAspLeL 39
                 :::      ||||||   ||||||.|.|   :::|||
     489 GAGCTCGGCTTCAAGCTGCCGTCGCCGGCGGACAAGATCGACCTCA 533

      39 euLeuMetPheValGlySerLeuGlyAlaIleValIleGlySerSerMet 55
         |||||::|||||||::|||||||||::|||||||||   |||:::
     534 GAGCTCATCTTCGTCGGCAGCCTCGGCGCCATCGTCATCGGCAGCAGCATG 583

      56 ProValPheLeuLeuPheGlyGlnMetValAsnGlyPheGlyLysAs 72
         |||||||||:::   :::|||:::::::::|||:::|||||||||:::
     584 CCGGTCTTCCTCCTCTTCGGCCAGCTCGTCAACGGCTTCGGCAAGGAC 633

      72 nGlnMetAsp 75
         :   |:|
     634 AAGATGGAC 642

seq_name: /tmp/ftp/cgi-bin/tmp/US09817762_2_COPY_1_76.rng.AAW60544

seq_documentation_block:
ID   AAW60544 standard; cDNA; 4827 BP
XX
AC   AAW60544;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Maize P-glycoprotein br2 cDNA.
XX
KW   Br2 gene; maize; corn; P-glycoprotein; transgenic plant;
KW   disease resistance; crop protection; dwarfing; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             91..4275
FT                   /*tag    a
XX
PN   WO200143819-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30821.
XX
PR   12-NOV-1999;  99US-0164886.
XX
PA   (BOHA/) BOHAL G S.
PA   (MOLI/) MOLIANI D S.
PA   (BRIG-) BRIGGS S P.
XX
PI   Bohal GS,  Moliani DS,  Briggs SP;
XX
DR   WPI; 2001-326088/34.
DR   P-PSDB; AAB20497.
XX
PT   Novel isolated P-glycoprotein polynucleotide, preferably br2 gene of
PT   maize, useful for modifying the growth of organisms, particularly
PT   plants and for enhancing resistance of plants to pathogens
XX
PS   Claim 1(c); Page 58-66; 73pp; English.
XX
CC   The present sequence is that of cDNA encoding maize P-glycoprotein
CC   BR2 (see AAB20497).  The cDNA was isolated from 10 day old B73
CC   seedlings in 4 overlapping portions by a combination of RT-PCR and
CC   5' RACE.  The br2 gene (located on maize chromosome 1L) was also
CC   isolated (see AAF30956).  Sequencing revealed extensive sequence
CC   and structural similarity with the multidrug resistance (MDR)-like
```

```
CC   an organism.  Such methods find use in agriculture, particularly in
CC   the breeding of dwarf crop plants.  Expression cassettes,
CC   transformed plants and their tissues, cells and seeds are also
CC   provided.  Brz nucleic acids also find use in the construction of
CC   expression vectors for subsequent transformation into plants, as
CC   probes for the isolation of other Polycoproteinlike genes, and as
CC   molecular markers.  Antisense sequences can be used to reduce the
CC   height of a plant.  The isolated gene provides environmentally safe
CC   alternatives to traditional methods of retarding stem elongation or
CC   growth, and an alternative to the use of synthetic pesticides in
CC   controlling plant pathogens.
XX
SQ   Sequence 6436 BP; 1671 A; 2476 C; 2284 G; 1748 T; 5 other

Alignment scores:
                Quality:  242.00              Length:      70
                  Ratio:  2.538                 Gaps:      1
 Percent Similarity:  74.286   Percent Identity: 41.429

Alignment_block:
 US-09-817-762-2_COPY_1_76 x AAF88488     .

 Align seq 1/1  to AAF88488  from 1  to 6336

        8 AspAlaLysIleValProAlaIleAlaIleGlyLysLysCysHisAsn 24
          :::|:|:::   :::   |||   |:|:::   :::|:|:::   :::
     1265 AGCCCAAAGATCGTTCCAGCTATAGCTATTGGTAAGAAGTGTCACAAC 1314

       25 nPro.....PheThrLysLeuPheSerPheAlaAspLysIleAspLys 39
            |||       :::     ||||||   ||||||||||   :::|||
     1315 NCCTGGTGGTCTTCAGTGCGACTCTTCTCCTTCGCTGACAAGATCGAC 1364

       40 cysLeuMetPheValGlySerLeuGlyAlaIleValIleGlyThrCysMet 55
          |:|||:::::::|||:::|||||||||:::||||||||||   |||:::
     1365 CGCTTCATGGTCACCCTCGGCGCCATCGTCATCGGCACCTGCATG 1411

       56 ProValPhePheLeuLeuPheGlyGlnMetValAsnGlyPheGlyLysAs 72
          |:|:|||||:::   :::|||||||||||||||:::|||||||||:::
     1412 CCTGTCTTCTTCCTGCTCTTCGGCCAGATGGTCAACGGCTTCGGCAAGAC 1461

       73 nGlnMetAsp 75
          :   |||
     1462 CCAGATGAC 1470

seq_name: /cgn2_6/ptodata/1/pcteus/US09_0817_0762_2_COPY_1_76.rng

seq_documentation_block:
ID   AAA41798 standard; cDNA; 896 BP
XX
AC   AAA41798;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEST.
XX
KW   human, mouse, xenopus, rat, secreted expressed sequence tag, sEST,
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antitumoral;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiulcer;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy,
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes, asthma, myeloid cell deficiency,
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease,
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200042196-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;  99WO-US24205.
XX
PR   15-OCT-1998;  98US-0104435.
XX
PA   (GENE ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 2000-317967/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders
XX
PS   Claim 1; Page 284; 618pp; English.
XX
CC   AAA41761 to AAA41893 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemokinetic, proliferative, immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic, antibacterial, antitumoral, antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic, antipsoriatic; antipyretic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used to gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA41760 to AAA41893 represent linker sequences which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence   99 BP; 23 A; 34 C; 99 G; 99 T; 0 other;

Alignment scores:
                Quality:  124.00              Length:      60
                  Ratio:  2.562                 Gaps:       0
 Percent Similarity:  80.000   Percent Identity: 66.667

Alignment_block:
 US-09-817-762-2_COPY_1_76 x AAA41798     .

 Align seq 1/1  to AAA41796  from 1  to  99

       17 GlnLysCysLysGlnGlnSerLeuProPheThrLysLeuPheSerPheAl 33
          :::|:|:::|||:::   ::::::       |||   :::|||   :::||
       1 ATGAAGAAGTGTCAGCAGTCCCTGCCCTTCACCAAGCTCTTCTCCTTCGC 50

       34 aAspLysIleAspLysCysMetPheValGlyCysLeuGlyAlaIleVal 50
          ||||   :::|||   |||   |||:::::::|||:::|||||||||||
       51 AGACTTCATCGACAAGTGCATGTTCGTCGGCTGCCTCGGCGCCATCGTC 100

       50 alHisGlySerSerMetProValPheThrLeuLeuThrGlyGlnMetVal 66
          :::|||||:::   :::|||:::::::::|||:::|||||||||:::
      101 ATCCACGGCTCCTCCATGCCCGTCTTCACCCTCCTCACCGGCCAGATGGTC 150

       67 AsnGlyPheGlyLysAsnGlnSerAsp 79
          :::::::|||   |||   :::   :::::
      151 AACGGCTTCGGCAAGAACCAGTCCGAC 177
```

```
[illegible] /NA [illegible] AA[illegible]

[illegible]
ID   AAZ4[illegible] standard; cDNA; [illegible] BP.
XX
AC   AAZ4[illegible]
XX
DT   [illegible] MAR [illegible] (first entry)
XX
DE   Murine multidrug resistance 1 (MDR 1) cDNA
XX
KW   Multidrug resistance; MDR [illegible] P-glycoprotein;
KW   transmembrane efflux pump; haematopoietic stem cell; transfection;
KW   bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW   gene therapy; gene replacement; genetic defect; thalassemia;
KW   [illegible] disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW   cytokines.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             [illegible]
FT                   /*tag= a
FT                   /product= "Murine MDR 1 protein"
XX
PN   W[illegible]-A1.
XX
PD   [illegible]
XX
PF   [illegible] MAY [illegible]; [illegible]
XX
PR   [illegible] MAY [illegible]; [illegible]
XX
PA   (STJU-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI   [illegible] K;
XX
DR   WPI; [illegible]
DR   P-PSDB; AAY[illegible].
XX
PT   Ex vivo expansion of haematopoietic stem cells transfected with a
PT   sequence encoding human multidrug resistance-1, used for bone marrow
PT   transplantation
XX
PS   [illegible] English.
XX
CC   This sequence represents cDNA encoding murine multidrug
CC   resistance protein MDR 1. MDR 1 is a transmembrane
CC   efflux pump, responsible for the export of drugs from cells,
CC   particularly cancer cells. The invention relates to transfecting
CC   haematopoietic stem cells with a vector encoding an MDR protein
CC   and [illegible] the modified cells. The modified haematopoietic stem
CC   cells are useful in bone marrow transplantation (to reconstitute
CC   haematopoietic systems in patients who have undergone chemotherapy or
CC   radiation therapy) and in ex vivo gene therapy of genetic defects in
CC   cells derived from haematopoietic stem cells, e.g. thalassaemia,
CC   Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC   cells can also be used to identify factors involved in regulating
CC   proliferation and differentiation in haematopoietic stem cells.
CC   Haematopoietic stem cells that express MDR 1 will be protected against
CC   chemotherapeutic agents, so can be engrafted while the patient is
CC   undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC   provides sufficient cells to permit standard biochemical analysis.
CC   Overexpression of MDR 1 allows cytokine driven expansion of
CC   haematopoietic stem cells by at least [illegible] fold compared with a maximum
CC   of 4 fold in known procedures.
XX
SQ   Sequence [illegible] BP; [illegible] A; [illegible] other;

[illegible]
          [illegible]               [illegible]
```

```
[illegible]

[illegible]
ID   AAA4[illegible]
XX
AC   AAA4[illegible]
XX
DT   [illegible] MAR [illegible]
XX
DE   Mouse [illegible]
XX
KW   Stem [illegible]
KW   [illegible]
KW   [illegible]
XX
OS   Mus musculus.
XX
FH   Key             [illegible]
FT   CDS             [illegible]
FT                   /*tag= [illegible]
XX
PN   W[illegible]
XX
PD   [illegible]
XX
PF   [illegible] MAY [illegible]
XX
PR   [illegible] MAY [illegible]
PR   [illegible] MAY [illegible]
XX
PA   (ST[illegible]) ST [illegible] HOSP[illegible]
XX
PI   [illegible]
XX
DR   WPI; [illegible]
DR   P-PSDB; AAB[illegible]
XX
PT   [illegible]
PT   [illegible]
PT   disorders, [illegible]
PT   [illegible]
XX
PS   [illegible]
XX
CC   The [illegible]
CC   [illegible]
CC   [illegible]
CC   [illegible]
CC   [illegible]
```

```
XX
PI   Brun KA, Ellens H, Yue L, Chenery RJ, Field JA;
XX
DR   WPI; 2001-[illegible]/[illegible].
DR   P-PSDB; AAY[illegible].
XX
PT   Isolated polynucleotide encoding a rat mdr1b2 drug transporter
PT   polypeptide, useful for treatment of e.g. cancer, autoimmune disease,
PT   central nervous system disorders
XX
PS   Claim 7; Column [illegible]; 17pp; English.
XX
CC   This sequence represents a polynucleotide encoding a 1275 amino acid
CC   rat mdr1b2 multispecific drug transporter polypeptide. The mdr1b2
CC   nucleotide sequence has cytostatic, antiinflammatory, cardiant,
CC   neuroprotective, immunosuppressive and nephrotropic activity.
CC   Understanding the functioning of the mdr1b2 polynucleotide and protein in
CC   transgenic animal models is useful for treating and preventing diseases
CC   such as cancer, inflammation, cardiovascular disease, central nervous
CC   system disorders, autoimmune disease, and kidney disease. The use of the
CC   protein in cell based, membrane based, or binding assays may enhance drug
CC   formulation, selection of formulation excipients and compound design.
XX
SQ   Sequence 4233 BP; 1216 A; 92[illegible] C; 1048 G; 1055 T; 0 other;


Alignment Scores:
                Quality:  122.00                  Length:      [illegible]
                  Ratio:  2.[illegible]                    Gaps:        0
  Percent Similarity:  80.[illegible]   Percent Identity:  [illegible].2[illegible]

Alignment Block:
 US-09-817-762-2_COPY_1_76 x AAZ[illegible]

 Align seq 1/1 to: AAZ[illegible] from: 1 to: 4233

        14 AlaGluAlaGlnLysLysLysGlnGlnSerLeuProPhePheLysLeuPh 30
           ::::::    ||||||:::|||||:::     ::::::    ::::||   ::::||
        89 AGTAAGAA[illegible]GAAGAAA[illegible]TATATT[illegible]CATCTT 138

        30 eSerPheAlaAspLysPheAspTyrLeuLeuMetPheValGlySerLeuG 47
           |   ::::|:|||    ::::||:|   :.|     ||    :::|||:::|||:
       [illegible] [illegible]AACTCTCG 188

        47 lyAlaIleValHisGlySerSerMetProValPhePheLeuLeuPheGly 63
           ::||:::::::| ||||:::    :::|||:::::::::::::|||::: ||:||
       189 [illegible]GA 238

        64 GluMetValAsnGlyPhe 69
              ||  ::::::::::|||
       239 [illegible]TACAGATAGTTTT 256

Seq name: [illegible]AAF27498

Seq documentation block:
ID   AAF27498 standard; cDNA; 4233 BP.
XX
AC   AAF27498;
XX
DT   25-APR-2001  (first entry)
XX
DE   Rat mdr1b2 multidrug resistance cDNA, SEQ ID NO: [illegible].
XX
KW   Rat; multidrug resistance; mdr1b2; multi drug transporter family;
KW   drug screening; pharmacokinetic analysis; oral absorption;
KW   formulation design; bioavailability; transgenic animal; knockout animal;
KW   inflammation; cardiovascular disease; central nervous system disorder;
KW   cancer; autoimmune disorder; kidney disease; ss.
XX
OS   Rattus sp.
XX
PN   US6[illegible]-B1.
XX
PD   02-JAN-200[illegible].
XX
PF   29-NOV-[illegible]
XX
PR   22-DEC-[illegible]
XX
PA   (SMIK ) SMI[illegible]
PA   (SMIK ) SMI[illegible]
XX
PI   Chenery RJ, [illegible] KA, Y[illegible]
XX
DR   WPI; [illegible]
DR   P-PSDB; AA[illegible]
XX
PT   New rat mu[illegible] protein [illegible]
PT   identifying [illegible] that are (ant)a[illegible]
PT   ratmdr1b2, [illegible] pounds with [illegible]
PT   characteriz[illegible]
XX
PS   Disclosure; [illegible]; English.
XX
CC   The invent[illegible] rat [illegible]
CC   protein, mdr1[illegible] (AAY[illegible]). Mdr1 spe[illegible]
CC   proteins are [illegible] which [illegible]
CC   intestinal epith[illegible], brain a[illegible]
CC   kidney epith[illegible] and liver hepat[illegible]
CC   by certain [illegible]. The rat mdr1b2 [illegible]
CC   screening [illegible] compounds that are [illegible]
CC   of mdr1b2 a[illegible] be used [illegible]
CC   predict ora[illegible] and pharmac[illegible]
CC   thus enables [illegible] formulation [illegible]
CC   compounds w[illegible] development [illegible]
CC   bioavailab[illegible]
CC   reduced var[illegible] and reduced [illegible]
CC   avoid inter[illegible] with human mdr1 [illegible]
CC   created using [illegible] the rat mdr1[illegible]
CC   insight int[illegible] and preventing human [illegible]
CC   inflammation, [illegible] disease, [illegible]
CC   autoimmune [illegible] and kidney disease[illegible]
CC   cDNA encoding [illegible]
XX
SQ   Sequence 4[illegible] A[illegible]


Alignment Scores:
                Quali[illegible]                  Length:
                  Rat[illegible]                    Gaps:
  Percent Similar[illegible]   Percent Identity:

Alignment Block:
 US-09-817-762-2_[illegible] x AAF274[illegible]

 Align seq 1/1 [illegible]

        14 AlaGluAla[illegible]SerLe[illegible]
        89 AGTAAAAA[illegible]

        30 eSerPheAla[illegible]TyrLeuLeuMet[illegible]
       139 [illegible]

        47 lyAlaIleVal[illegible]SerMetPr[illegible]
       189 [illegible]

        64 GluMetVa[illegible]
       239 TACATAGA[illegible]

Seq name: [illegible]

Seq documentation [illegible]
```

```
[illegible] AA[illegible] standard; cDNA; [illegible] BP.
XX
AC   AA[illegible]
XX
DT   14 MAR [illegible] (first entry)
XX
DE   Human wild type multidrug resistance 1 (MDR 1) cDNA.
XX
KW   Multidrug resistance; MDR 1; P glycoprotein;
KW   transmembrane efflux pump; haematopoietic stem cell; transduction;
KW   bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW   gene therapy; gene replacement; genetic defect; thalassaemia;
KW   [illegible]'s disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW   cytokine; wild type; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3843
FT                   /*tag= a
FT                   /product= "Human wild type MDR 1 protein"
FT   mutation        replace([illegible], [illegible])
FT                   /*tag= b
FT                   /note= "cDNA sequence of G185V human mutant MDR 1 given
FT                   in AA[illegible]"
XX
PN   W[illegible] A[illegible]
XX
PD   [illegible]
XX
PF   27 MAY [illegible];   [illegible]WO-US11[illegible].
XX
PR   28 MAY 1998;   98US-0086988.
XX
PA   (STJU-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI   Sorrentino B[illegible], [illegible] K;
XX
DR   WPI; [illegible]/[illegible].
DR   P-PSDB; AAY[illegible].
XX
PT   Ex vivo expansion of hematopoietic stem cells transduced with a
PT   sequence encoding human multidrug resistance 1, used for bone marrow
PT   transplantation
XX
PS   Claim 1; Page 68-70; 113pp; English.
XX
CC   This sequence represents cDNA encoding human wild type
CC   multidrug resistance protein MDR 1. MDR-1 is a transmembrane
CC   efflux pump, responsible for the export of drugs from cells,
CC   particularly cancer cells. Wild type MDR-1 shows increased
CC   resistance to etoposide and decreased resistance to vinca
CC   alkaloids compared with a mutant form (AAY[illegible]) where the Gly at
CC   position 185 is replaced by Val. The invention relates to transducing
CC   haematopoietic stem cells with nucleic acid encoding an MDR protein
CC   and culturing the modified cells. The modified haematopoietic stem
CC   cells are useful in bone marrow transplantation (to reconstitute
CC   haematopoietic systems in patients who have undergone chemotherapy or
CC   radiation therapy) and in ex vivo gene therapy of genetic defects in
CC   cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC   Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC   cells can also be used to identify factors involved in regulating
CC   proliferation and differentiation in haematopoietic stem cells.
CC   Haematopoietic stem cells that express MDR 1 will be protected against
CC   chemotherapeutic agents, so can be engrafted while the patient is
CC   undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC   provides sufficient cells to permit standard biochemical analysis.
CC   Overexpression of MDR 1 allows cytokine-driven expansion of
CC   haematopoietic stem cells by at least 10-fold compared with a maximum
CC   of 4-fold of known procedures
XX
SQ   Sequence [illegible] BP; [illegible] A; [illegible] C; [illegible] G; [illegible] T; [illegible] other;

Alignment Scores:
Pred. No.:       [illegible]          Length:      [illegible]
Score:           [illegible]          Matches:     [illegible]
Ratio:           [illegible]          Gaps:        [illegible]
Percent Similarity: [illegible]   Percent Identity: [illegible]

Alignment Block:
US-09-817-762-2_COPY_1_76 x AAZ4[illegible]

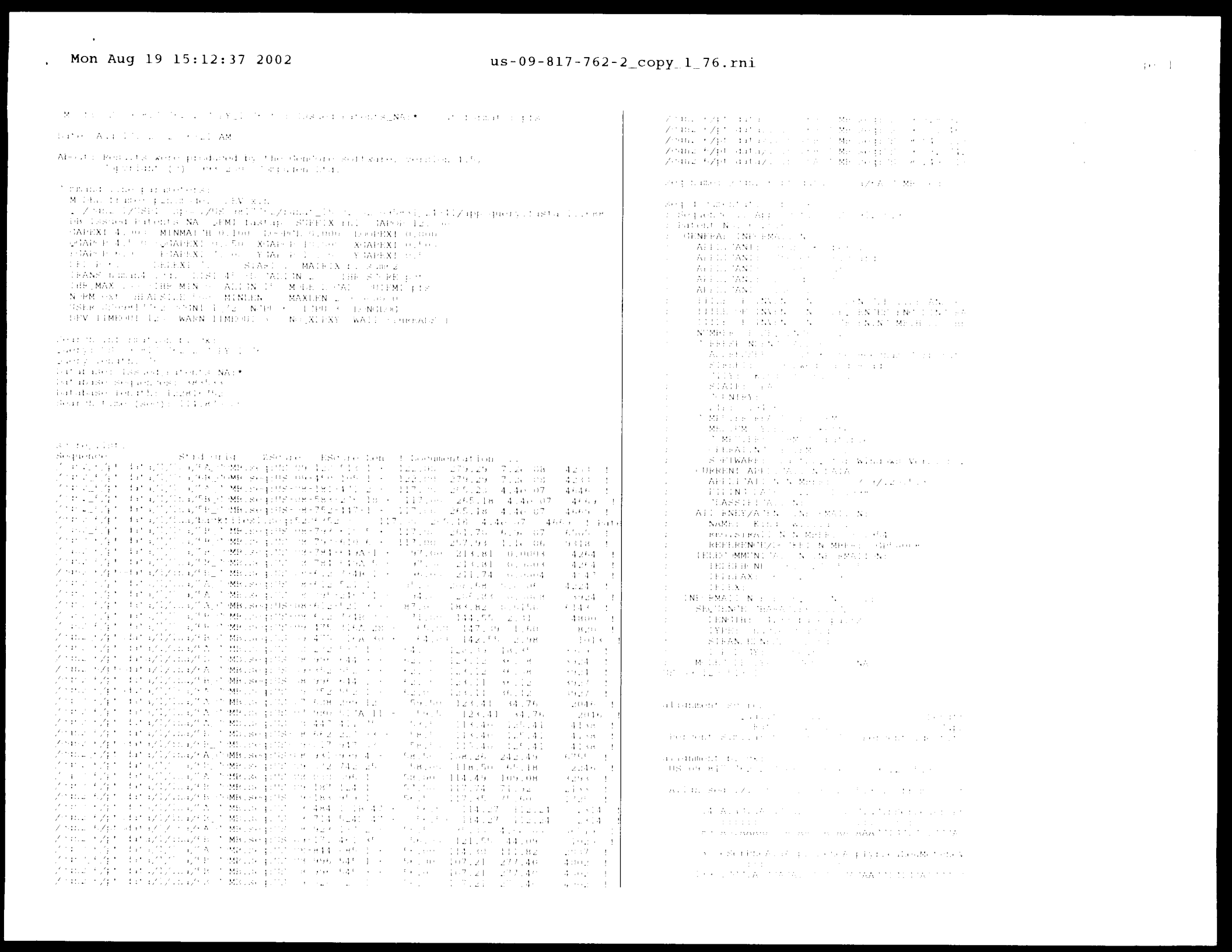
Align seq 1/1 to: AAZ4[illegible] from: 1 to: 3860

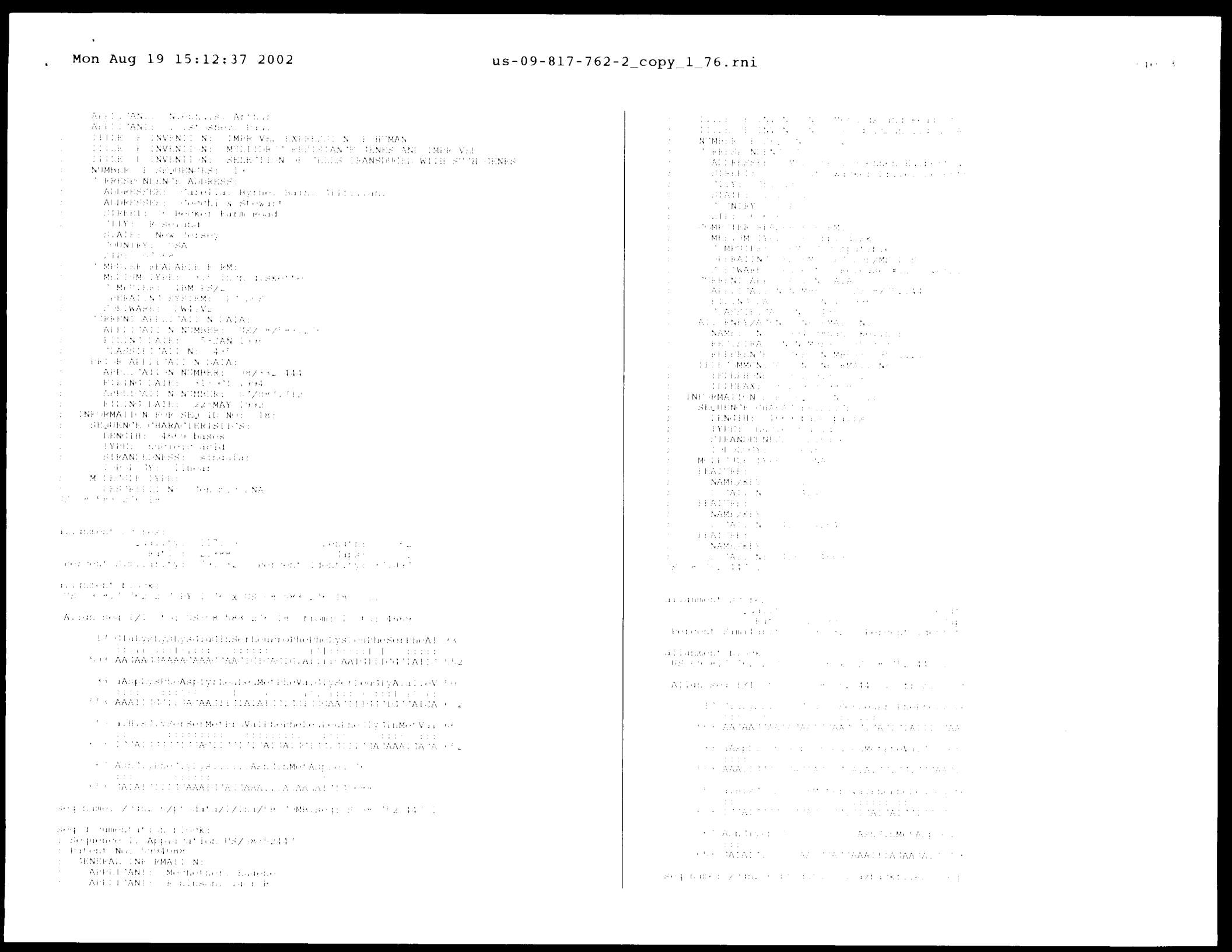
    17 [illegible]LysLysLysIleGlnSerLeuProPhePheLysLeuPheSerPheAl[illegible] 33
       [illegible]
    [illegible] AA[illegible]AAAAGAAA[illegible]AA[illegible] 128

    [illegible] aAsp[illegible]MetPheVal[illegible]Gly[illegible] 50
       [illegible]
   129 AAA[illegible] 178

    50 aHisGlySerSerMetProValPhePheLeuLeuPheGlyGlnMetVal 66
       [illegible]
   179 [illegible]GGAGAAA[illegible]A 228

    67 AsnGlyPheGlyLys......AsnGlnMetAsp[illegible] 76
       [illegible]
   229 [illegible]AAA[illegible]GAAAT[illegible]AGAAGA[illegible] 254
```





```
seq_documentation_block:
; Patent No. 5676944
;    APPLICANT: Robinson, Eric B.;Pastan, Ira H.;Gottesman,
;Michael M.
;    TITLE OF INVENTION: A METHOD FOR PRODUCING A SECRETED
;SEQUENCE AND FACILITATING PURIFICATION OF A DESIRED PROTEIN
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: 08/207,959A
;      FILING DATE: 24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 992,535
;      FILING DATE: 21-AUG-1992
;      APPLICATION NUMBER: 937,610
;      FILING DATE: 28-MAR-1995
;SEQ ID NO: 2:
;      LENGTH: 1659
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
;    FRAGMENT TYPE: internal
;    ORIGINAL SOURCE:
;      ORGANISM: Pseudomonas exotoxin
;    FEATURE:
;      NAME/KEY: CDS
;      LOCATION: 1..1659
;      IDENTIFICATION METHOD: experimental
;      OTHER INFORMATION: /product= "PE40"
;      /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      AUTHORS: Pastan, I.H.
;      AUTHORS: Fitzgerald, D.J.
;      AUTHORS: Chaudhary, V.K.
;      AUTHORS: Gallo, M.G.
;      JOURNAL: Nature
;      VOLUME: 313
;      PAGES: 144-147
;      DATE: 1985
;    PUBLICATION INFORMATION:
;      AUTHORS: Kondo, T.
;      AUTHORS: Hwang, J.
;      AUTHORS: Adhya, S.
;      JOURNAL: J. Biol. Chem.
;      VOLUME: 263
;      PAGES: 16530-16535
;      DATE: 1988
;    PUBLICATION INFORMATION:
;      AUTHORS: Seetharam, R.
;      AUTHORS: Sharma, S.K.
;      JOURNAL: J. Biol. Chem.
;      VOLUME: 266
;      PAGES: 17376-17381
;      DATE: 1991
;    PUBLICATION INFORMATION:
;      AUTHORS: Siegall, C.B.
;      AUTHORS: Chaudhary, V.K.
;      AUTHORS: FitzGerald, D.J.
;      AUTHORS: Pastan, I.
;      JOURNAL: Cancer Res.
;      VOLUME: 49
;      PAGES: 3943-3948
;      DATE: 1989
;    PUBLICATION INFORMATION:
;      AUTHORS: Chaudhary, V.K.
;      AUTHORS: Jinno, Y.
;      AUTHORS: FitzGerald, D.J.
;      AUTHORS: Kondo, T.
;      AUTHORS: Gallo, M.G.
;      AUTHORS: Uchida, T.
;      AUTHORS: Pastan, I.
;      JOURNAL: Proc. Natl. Acad. Sci. USA
;      VOLUME: 86
;      PAGES: 1026-1030
;      DATE: 1989
;    PUBLICATION INFORMATION:
;      AUTHORS: Kondo, T.
;      AUTHORS: Hwang, J.
;      AUTHORS: Adhya, S.
;      JOURNAL: J. Biol. Chem.
;      VOLUME: 263
;      PAGES: 16530-16535
;      DATE: 1988
;    PUBLICATION INFORMATION:
;      AUTHORS: Seetharam, R.
;      AUTHORS: Sharma, S.K.
;      JOURNAL: J. Biol. Chem.
;      VOLUME: 266
;      PAGES: 17376-17381
;      DATE: 1991
;    PUBLICATION INFORMATION:
;      AUTHORS: Siegall, C.B.
;      AUTHORS: Chaudhary, V.K.
;      AUTHORS: FitzGerald, D.J.
;      AUTHORS: Pastan, I.
;      JOURNAL: Cancer Res.
;      VOLUME: 49
;      PAGES: 3943-3948
;      DATE: 1989
;    PUBLICATION INFORMATION:
;      AUTHORS: Chaudhary, V.K.
;      AUTHORS: Jinno, Y.
;      AUTHORS: FitzGerald, D.J.
;      AUTHORS: Kondo, T.
;      AUTHORS: Gallo, M.G.
;      AUTHORS: Uchida, T.
;      AUTHORS: Pastan, I.
;      JOURNAL: Proc. Natl. Acad. Sci. USA
;      VOLUME: 86
;      PAGES: 1026-1030
;      DATE: 1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 08-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 07-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 10-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Berman, Richard J.
;      REGISTRATION NUMBER: 39,105
;      REFERENCE/DOCKET NUMBER: P1614 7007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (202)639-5000
;      TELEFAX: (202)639-4810
;  INFORMATION FOR SEQ ID NO: 5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 6505 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: circular
;    MOLECULE TYPE: DNA
US-08-794-610-5


Alignment scores:
                Quality:  117.00              Length:      62
                  Ratio:  2.088                 Gaps:       1
  Percent Similarity:  79.032   Percent Identity:  37.097

Alignment block:
US-08-794-610-5  (1) to (6505)

 Align seq 1/1  to: 6505  from: 1  to: 6505

       17 GlnLysLysLysGlnGlnSerLeuProPheIleLysLeuPheSerPheA1 44
          :::|:::|:|:|::  ::::::     |||:::::|||   :::::
     1645 AACAAAAAAAAAAAAAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1694

       45 aAspIlePheAspIleAspMetHisValSerIleGlyAlaIleV 90
          ::::  :::|||  |||  |||  |||||::|||:|||||||
     1695 AAAGATTTTGATATAGACATGCACGTAAGCATCGGTGCAATCGA 1739

       91 alHisGlyGlyThrGlnAlaIleLysProIleAlaGlnSerIleA 136
          ||||||||||||||||||||||||||||||||||||||||||||
     1740 CGTGCACGGTGGAACGCAGGCAATCAAACCAATCGCACAGTCAATC 1784

      137 AspGlyArgLeuAlaAlaValIleGlyThrMetThrAlaGlu 176
          :::  |||::::::  |||  |||||
     1785 GACGGACGCCTGGCAGCAGTGATCGGTACAATGACCGCAGAA 1824

      177 AsnGlnMetAsp 188
          ||||| |||||
     1825 AACCAAATGGAT 1836

seq_documentation_block:
; Sequence 5, Application US/08793610
; Patent No. 5676744
;  GENERAL INFORMATION:
;    APPLICANT: BAUM, Christopher
;    APPLICANT: STERKING HARBERS, Carol
;    APPLICANT: OSTERTAG, Wolfram
;    TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
;    TITLE OF INVENTION: FOR GENE TRANSFER
;    NUMBER OF SEQUENCES: 6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
;      STREET: 655 Fifteenth Street N.W. Suite 330
;      CITY: Washington
;      STATE: D.C.
;      COUNTRY: U.S.A.
;      ZIP: 20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/793,610
;      FILING DATE: 07-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 08-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 07-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/207,959A
;      FILING DATE: 10-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Berman, Richard J.
;      REGISTRATION NUMBER: 39,105
;      REFERENCE/DOCKET NUMBER: P1614 7007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (202)639-5000
;      TELEFAX: (202)639-4810
;  INFORMATION FOR SEQ ID NO: 5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 6505 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: circular
;    MOLECULE TYPE: DNA
US-08-794-610-5


Alignment scores:
                Quality:  117.00              Length:      62
                  Ratio:  2.088                 Gaps:       1
  Percent Similarity:  79.032   Percent Identity:  37.097

Alignment block:
US-08-794-610-5  (1) to (6505)

 Align seq 1/1  to: 6505  from: 1  to: 6505

       17 GlnLysLysLysGlnGlnSerLeuProPheIleLysLeuPheSerPheA1 44
          :::|:::|:|:|::  ::::::     |||:::::|||   :::::
     1645 AACAAAAAAAAAAAAAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1694

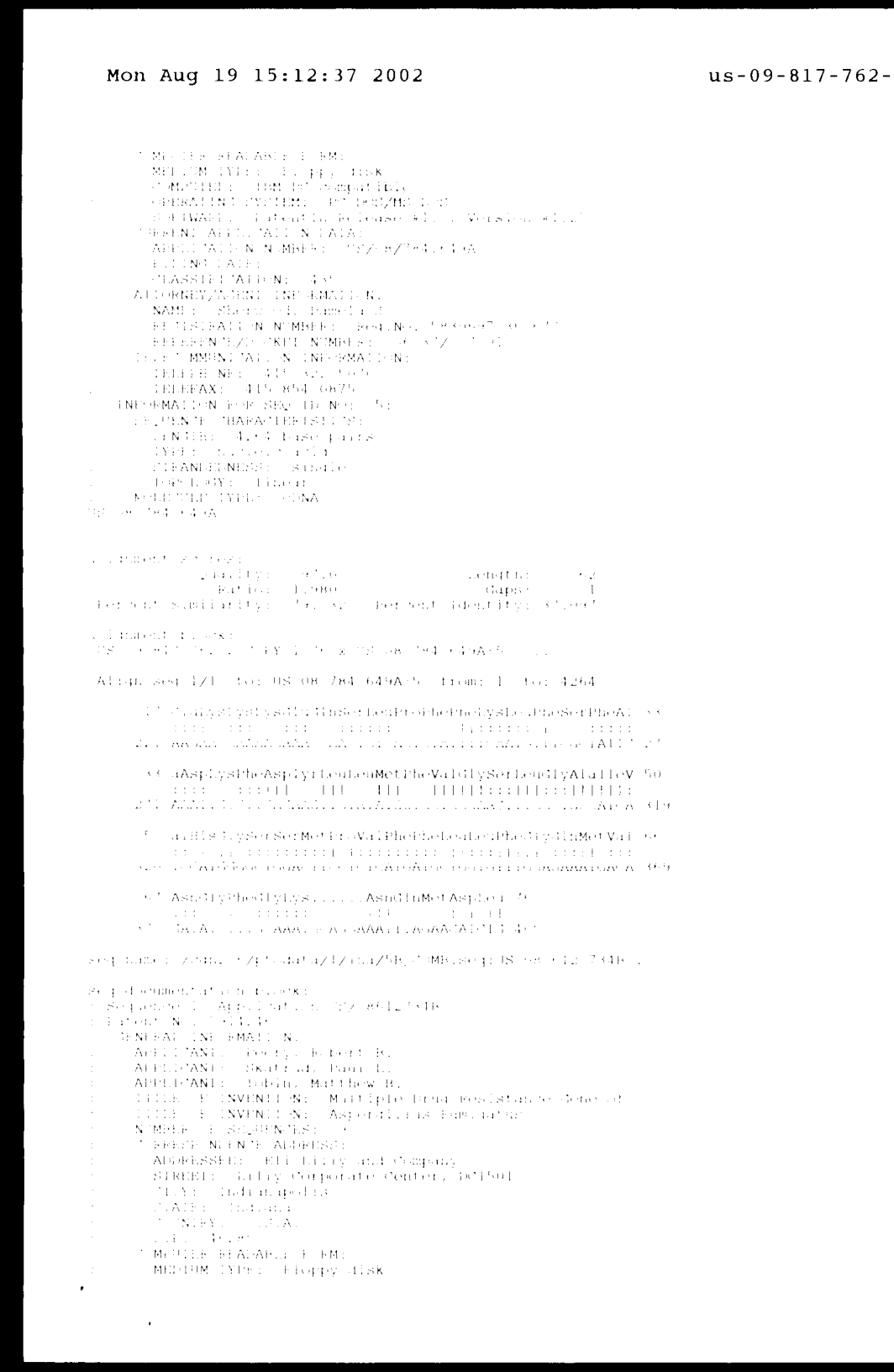
       45 aAspIlePheAspIleAspMetHisValSerIleGlyAlaIleV 90
          ::::  :::|||  |||  |||  |||||::|||:|||||||
     1695 AAAGATTTTGATATAGACATGCACGTAAGCATCGGTGCAATCGA 1739

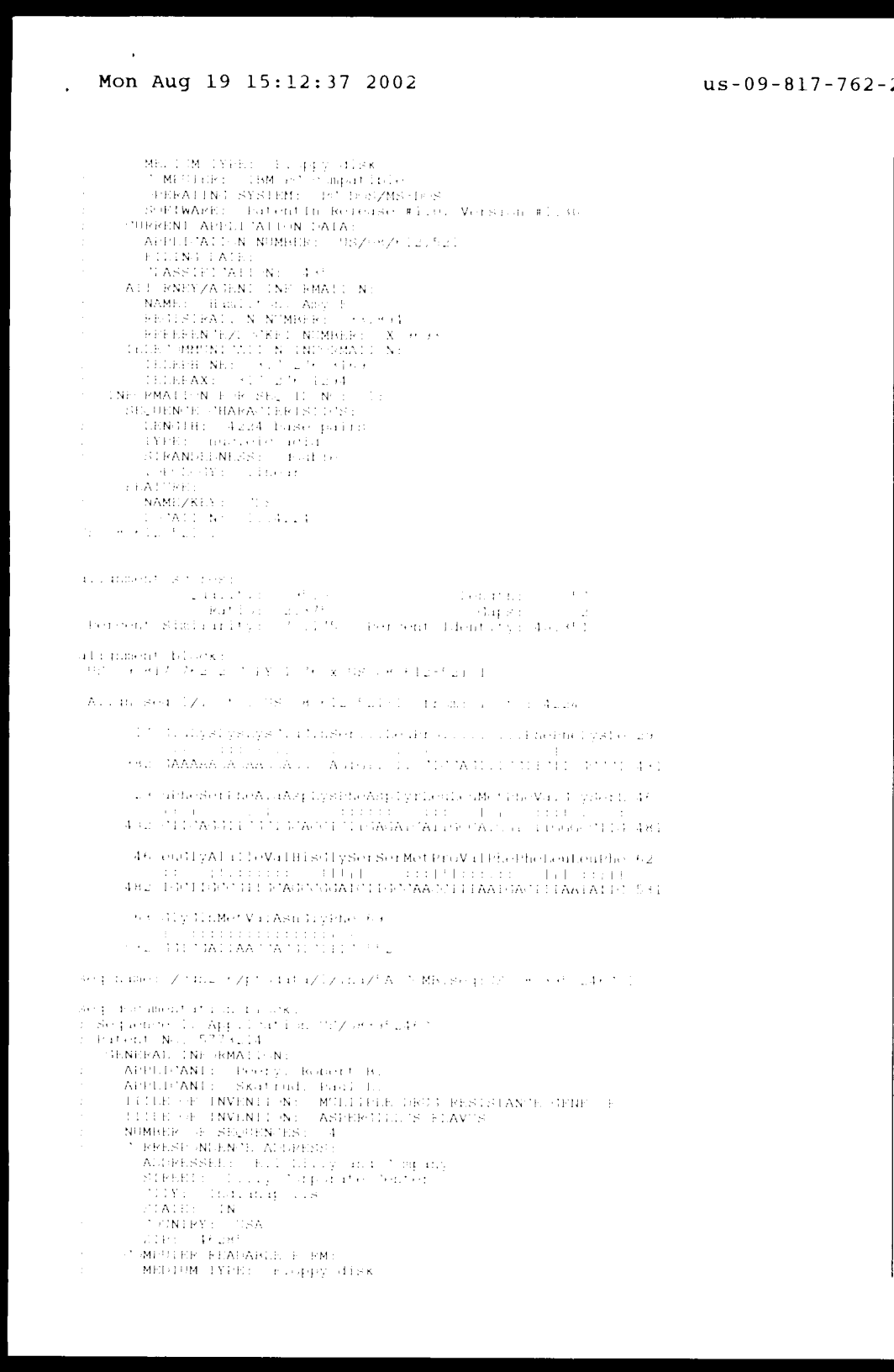
       91 alHisGlyGlyThrGlnAlaIleLysProIleAlaGlnSerIleA 136
          ||||||||||||||||||||||||||||||||||||||||||||
     1740 CGTGCACGGTGGAACGCAGGCAATCAAACCAATCGCACAGTCAATC 1784

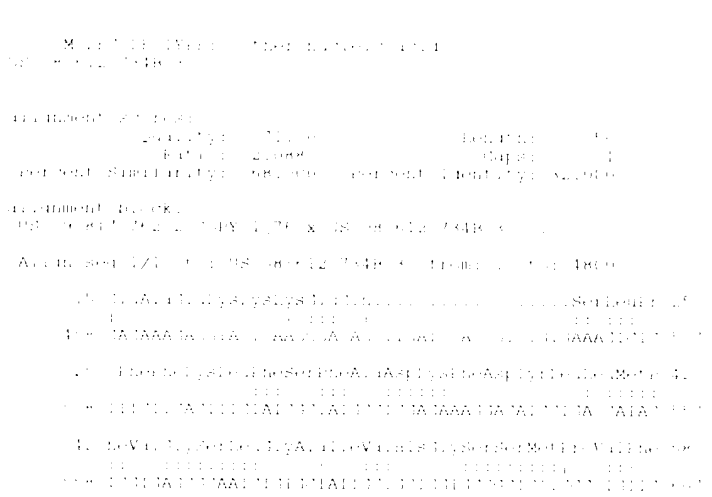
      137 AspGlyArgLeuAlaAlaValIleGlyThrMetThrAlaGlu 176
          :::  |||::::::  |||  |||||
     1785 GACGGACGCCTGGCAGCAGTGATCGGTACAATGACCGCAGAA 1824

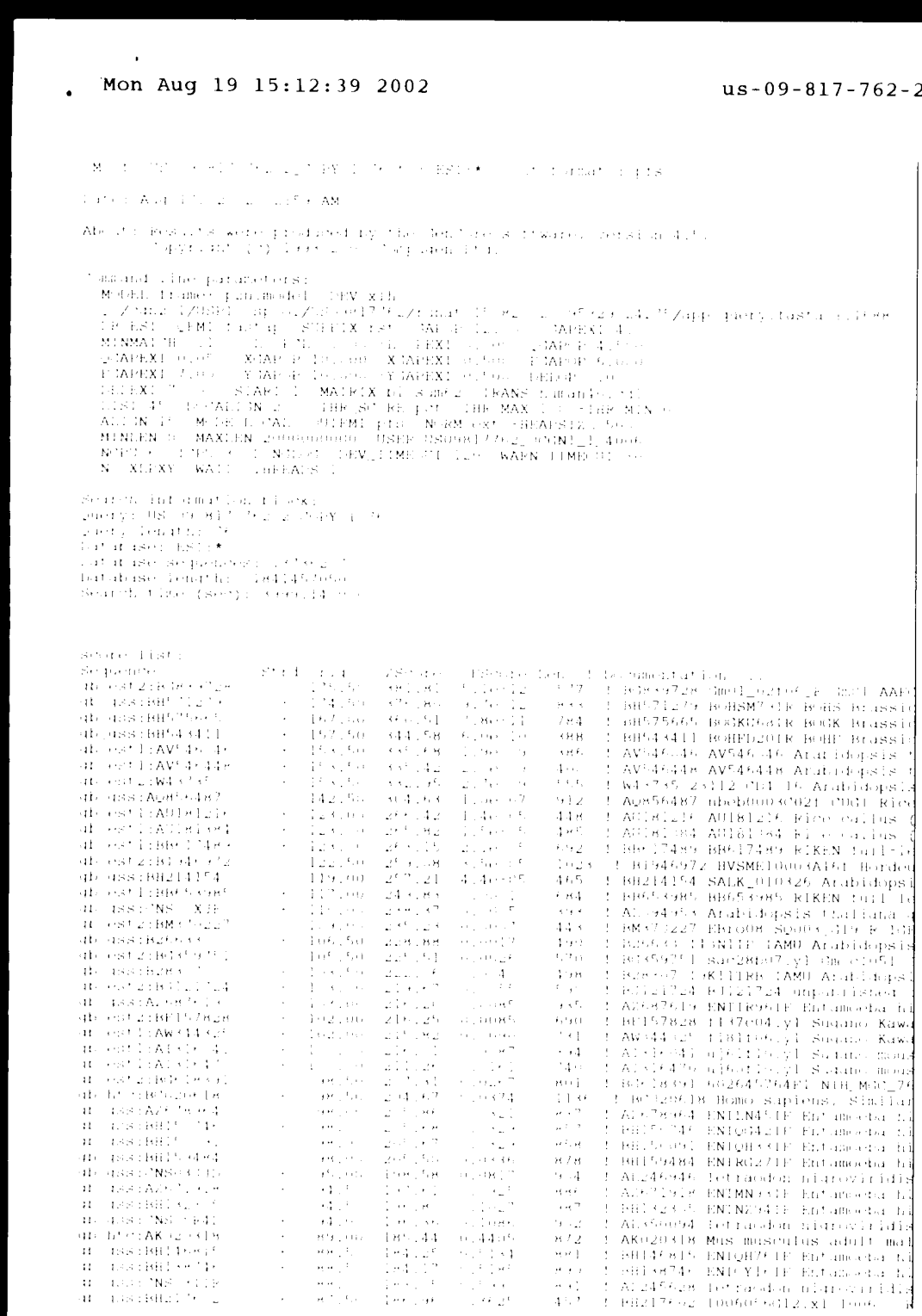
      177 AsnGlnMetAsp 188
          ||||| |||||
     1825 AACCAAATGGAT 1836

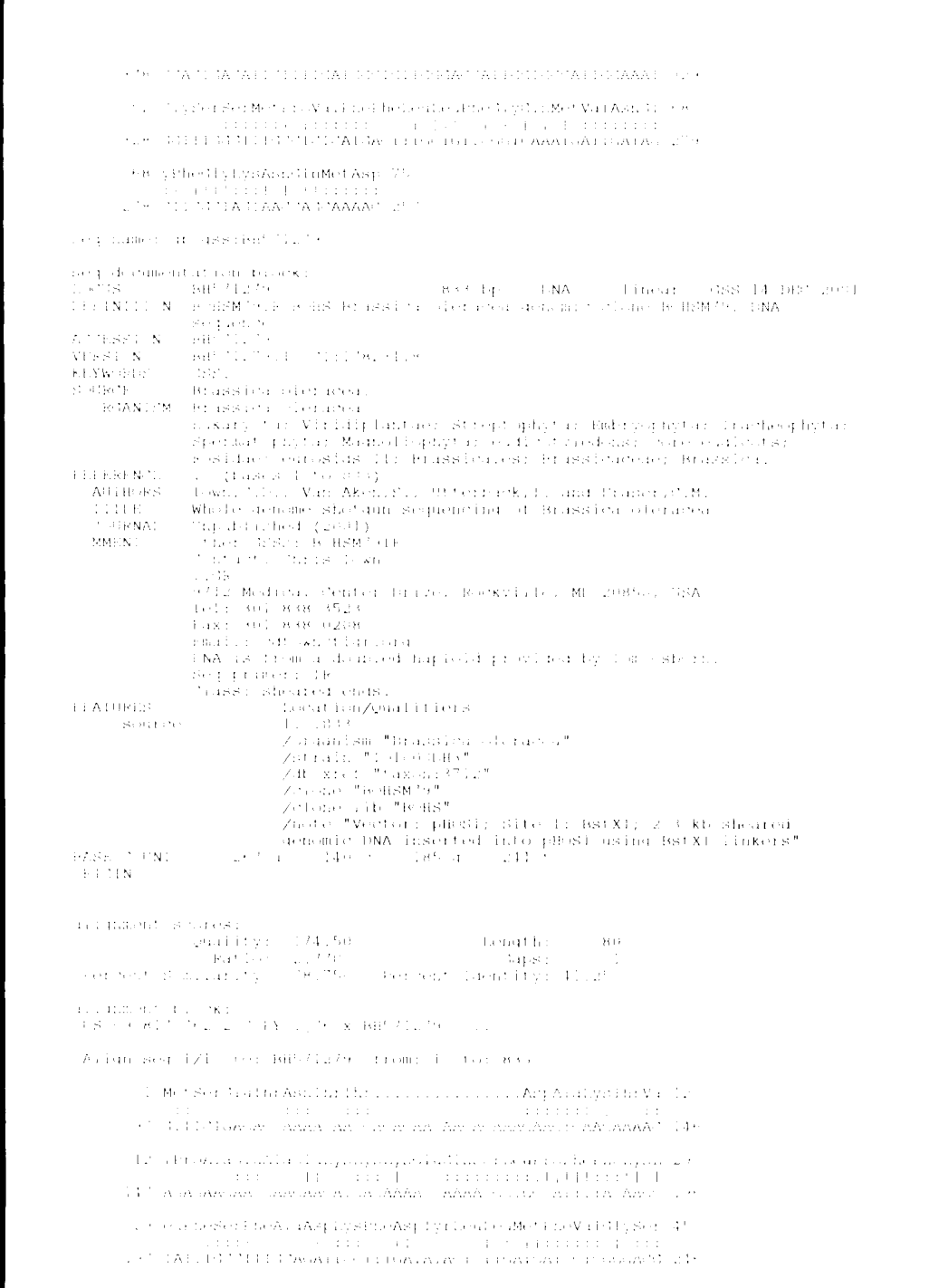
seq_documentation_block:
; Sequence 5, Application US/08793610
; Patent No. 5676744
;  GENERAL INFORMATION:
;    APPLICANT: BAUM, Christopher
;    APPLICANT: STERKING HARBERS, Carol
;    APPLICANT: OSTERTAG, Wolfram
;    TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
;    TITLE OF INVENTION: FOR GENE TRANSFER
;    NUMBER OF SEQUENCES: 6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
;      STREET: 655 Fifteenth Street N.W. Suite 330
;      CITY: Washington
;      STATE: D.C.
;      COUNTRY: U.S.A.
;      ZIP: 20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
```

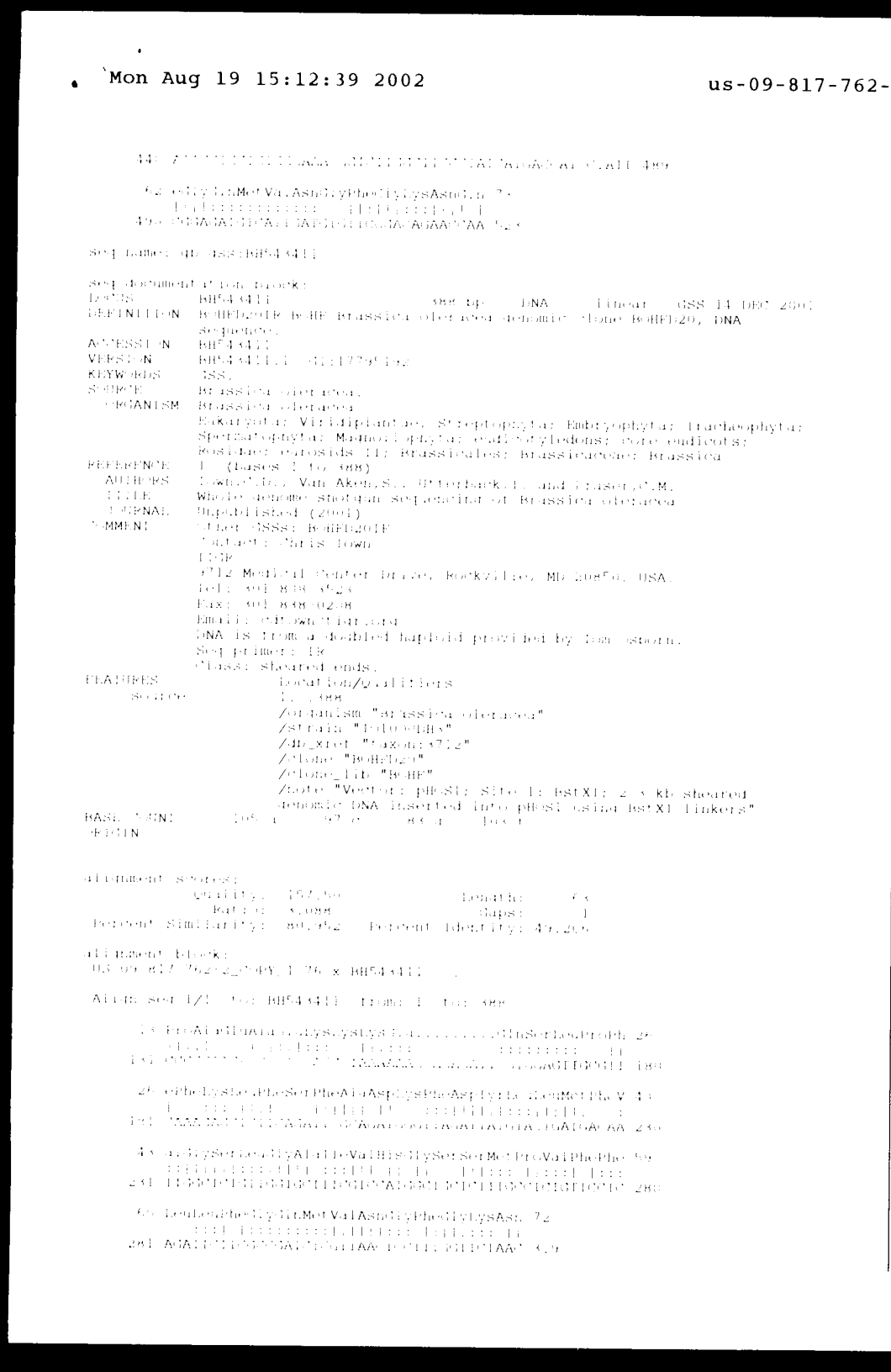












Seq_feature: misc_feature 1..488

Seq_document:Patent Data Block:
LOCUS BH543411 488 bp DNA linear GSS 14-DEC-2001
DEFINITION BcHEH2oIF BcHE Brassica oleracea genomic clone BcHEH20, DNA sequence.
ACCESSION BH543411
VERSION BH543411.1 GI:17779162
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Cronosids II; Brassicales; Brassicaceae; Brassica
REFERENCE 1 (bases 1 to 488)
AUTHORS Lowe,C.J., Van Aken,S., Osterbaken,I. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Clone ESSS: BcHEH20IF
Contact: Chris Lowe
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301 838 3623
Fax: 301 838 0208
Email: clowe@tigr.org
DNA is from a doubled haploid provided by Ian Storm.
Seq primer: IR
Class: Sheared ends.
FEATURES Location/Qualifiers
source 1..488
/organism="Brassica oleracea"
/strain="DH1 CBHs"
/db_xref="taxon:3712"
/clone="BcHEH20"
/clone_lib="BcHE"
/note="Vector: pBeST; Site 1: BstXI; 2 x kb sheared genomic DNA inserted into pBeST using BstXI linkers"
BASE COUNT 106 a 92 c 88 g 102 t
ORIGIN

alignment scores:
Quality: 152.50 Length: 488
Ratio: 8.088 Gaps: 1
Percent Similarity: 80.962 Percent Identity: 47.200

alignment block:
US-09-817-762-2_COPY_1_76 x BH543411 .

Align Seq 1/1 to: BH543411 from: 1 to: 488

Seq_feature: misc_feature 1..488

Seq_document:Patent Data Block:
LOCUS AV543411 488 bp DNA linear GSS 14-DEC-2001
DEFINITION AV543411 AvHEH2oIF AvHE Arabidopsis thaliana genomic clone AvHEH20, DNA sequence.
ACCESSION AV543411
VERSION AV543411.1 GI:17779162
KEYWORDS GSS.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Cronosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE 1 (bases 1 to 488)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale structural analysis of cDNAs in a normalized directional library from Arabidopsis thaliana
JOURNAL DNA Res. 7 (2), 175-180 (2000)
MEDLINE 20341141
COMMENT Contact: Asamizu Erika
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp
FEATURES
source 1..488
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RAFL21-17-P21"
/clone_lib="Arabidopsis thaliana Full-length cDNA Library"
/note="Vector: pBS; Site 1: EcoRI; Site 2: NotI; cDNA was inserted into pBS using EcoRI and NotI linkers"
BASE COUNT
ORIGIN

alignment scores:
Quality:
Ratio:
Percent Similarity: Percent Identity:

alignment blocks:
US-09-817-762-2_COPY_1_76 x AV543411 .

Align Seq 1/1 to: AV543411 from: 1 to: 488

Seq_feature: misc_feature 1..488

Seq_document:Patent Data Block:
LOCUS AV543411 488 bp DNA linear GSS 14-DEC-2001
DEFINITION AV543411 AvHEH2oIF AvHE Arabidopsis thaliana genomic clone AvHEH20, DNA sequence.
ACCESSION AV543411

```
VERSION     AV434448.1  GI:6478...
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale structural analysis of cDNAs in Arabidopsis thaliana genome
            1: 2,811,611 bp genomic DNA sequence
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20377475
COMMENT     On Jan 7, 2008 this sequence version replaced gi:6478...
            Contact: Thomas Nowak
            MSU DOE Plant Research Laboratory
            Michigan State University
            MSU DOE PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,MI
            Tel: 517 ... 0854
            Fax: 517 ... 9168
            Email: ...@msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1..498
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="H4E017"
                     /clone_lib="CD4-10"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day old"
                     /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
                     BamHI; Using 5 ug of polyadenylated mRNA from 3 day old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo-d(T) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia). Second strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted. After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end. The cDNA was purified from unligated
                     adapters by spun column chromatography using Sephacryl
                     S-400 and size fractionated on a 1% low melting point
                     gel. Size selected cDNAs (> 0.6 Kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:chloroform extracted and precipitated using 2.5 M
                     NH4OAc (pH 7)/ethanol. A portion of each cDNA
                     size fraction (0.1 ng) was co-precipitated with 1 ng of
                     LambdaZapII (Stratagene/EcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene). We have determined that
                     although first strand cDNA synthesis was initiated using
                     oligo-d(T), almost all of the cDNAs begin 8-10 bp from the poly A
                     tail. The reason for the loss of the poly A tail is not
                     known, but is expected to occur during the Klenow repair of nicked ends before the
                     addition of linkers (5' 5' exo instead of 5' 3' pol).
                     When this library is used please reference the ABRC and
                     Kieber, J. et al. (1993) Cell 72, 427-441."
BASE COUNT      ... a ... c ... g ... t ... others
ORIGIN


Alignment Scores:
                Frame:  ...
                Query:  ...                 Length:  ...
                  Score:  ...                   Gaps:  ...
 Percent Similarity:  80.952%   Percent Identity:  49.206%

Alignment Blocks:
 Sequence 1 of 498 vs AV434448.1 from: 1 to: 498

 Alignment of: 1 vs AV434448.1 from: 1 to: 498

      1: MetValProAsnThrAlaAspLysLysSerGlu...GluSerLeuArgGln 26
         |:|||  ||||||  ||||||  ||  |||||||||  |
      1: ATGGTACCAAATACTGCAGACAAAAAGAGTGAAGAGTCTCTTCGTCAG 47

     27: GlnGlyLeuGlnSerPheAsnAspLysPheAspLysTyrLeuGluMetPheVal 43
         |  |||||||  |||||||  |||||||  |||||||  |||||||||||
     48: TAAAGCCTTAATCAATCTTGGAAAGATAAATCTTTAGATCAATTCGTTGAA 97

     44: AsnLysSerLeuLysAsnValAlaSerSerProSerSerMetGlnValMetVal 60
         ||  |||  ||||||||  ||||||||  ||||||  ||||||||  |||||
     98: TACGTTCTTGAGAACGTTGCTAGTTCTCCTTCTTCTATGCAAGTTATGGTT 147

     61: SerSerProGlnAspGlnGlnProGlnGlnGlnAlaGlnGlnProGlnGlnPro 77
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    148: TCTTCTCCTCAAGATCAACAACCTCAACAACAAGCTCAACAACCTCAACAACCT 197

         ...
```

```
VERSION     AV434448.1  GI:6478...
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale structural analysis of cDNAs in Arabidopsis thaliana genome
            1: 2,811,611 bp genomic DNA sequence
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20377475
COMMENT     On Jan 7, 2008 this sequence version replaced gi:6478...
            Contact: Thomas Nowak
            MSU DOE Plant Research Laboratory
            Michigan State University
            MSU DOE PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,MI
            Tel: 517 ... 0854
            Fax: 517 ... 9168
            Email: ...@msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1..498
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="H4E017"
                     /clone_lib="CD4-10"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day old"
                     /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
                     BamHI; Using 5 ug of polyadenylated mRNA from 3 day old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo-d(T) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia). Second strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted. After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end. The cDNA was purified from unligated
                     adapters by spun column chromatography using Sephacryl
                     S-400 and size fractionated on a 1% low melting point
                     gel. Size selected cDNAs (> 0.6 Kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:chloroform extracted and precipitated using 2.5 M
                     NH4OAc (pH 7)/ethanol. A portion of each cDNA
                     size fraction (0.1 ng) was co-precipitated with 1 ng of
                     LambdaZapII (Stratagene/EcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene). We have determined that
                     although first strand cDNA synthesis was initiated using
                     oligo-d(T), almost all of the cDNAs begin 8-10 bp from the poly A
                     tail. The reason for the loss of the poly A tail is not
                     known, but is expected to occur during the Klenow repair of nicked ends before the
                     addition of linkers (5' 5' exo instead of 5' 3' pol).
                     When this library is used please reference the ABRC and
                     Kieber, J. et al. (1993) Cell 72, 427-441."
BASE COUNT      ... a ... c ... g ... t ... others
ORIGIN


Alignment Scores:
                Frame:  ...
                Query:  ...                 Length:  ...
                  Score:  ...                   Gaps:  ...
 Percent Similarity:  80.952%   Percent Identity:  49.206%

Alignment Blocks:
 Sequence 1 of 498 vs WI7735 from: 1 to: 498

 Alignment of: 1 vs WI7735 from: 1 to: 498

      1: MetValProAsnThrAlaAspLysLysSerGlu...GluSerLeuArgGln 26
         |:|||  ||||||  ||||||  ||  |||||||||  |
      1: ATGGTACCAAATACTGCAGACAAAAAGAGTGAAGAGTCTCTTCGTCAG 47

     27: GlnGlyLeuGlnSerPheAsnAspLysPheAspLysTyrLeuGluMetPheVal 43
         |  |||||||  |||||||  |||||||  |||||||  |||||||||||
     48: TAAAGCCTTAATCAATCTTGGAAAGATAAATCTTTAGATCAATTCGTTGAA 97

     44: AsnLysSerLeuLysAsnValAlaSerSerProSerSerMetGlnValMetVal 60
         ||  |||  ||||||||  ||||||||  ||||||  ||||||||  |||||
     98: TACGTTCTTGAGAACGTTGCTAGTTCTCCTTCTTCTATGCAAGTTATGGTT 147

     61: SerSerProGlnAspGlnGlnProGlnGlnGlnAlaGlnGlnProGlnGlnPro 77
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    148: TCTTCTCCTCAAGATCAACAACCTCAACAACAAGCTCAACAACCTCAACAACCT 197

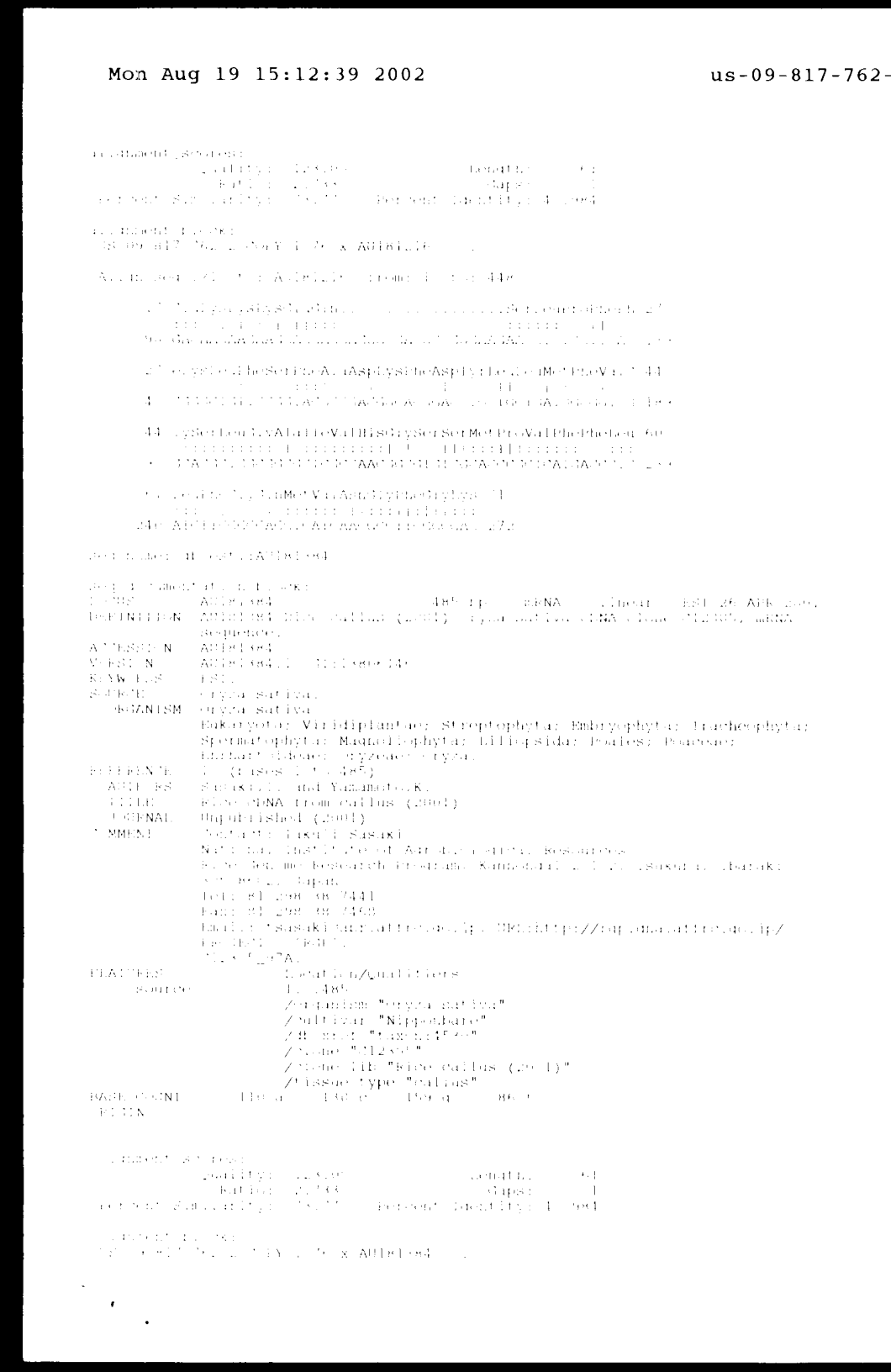
         ...
```

```
LOCUS       [illegible]
DEFINITION  [illegible] BAC Rice BAC Library (Boo01) Oryza sativa genomic
            [illegible] DNA sequence.
ACCESSION   [illegible]
VERSION     [illegible]
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Wing,R.A. and Soderlund,C.A.
  TITLE     A BAC End Sequence Survey of Rice ... Sequence to Discoveries
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            [illegible] Bldg., Clemson, SC [illegible], USA
            Tel: 864 656 3808
            Fax: 864 656 4105
            Email: rwing@clemson.edu
            Polished: [illegible]
            Class: HTG
            High quality sequence start: 60
            High quality sequence stop: 212
FEATURES             Location/Qualifiers
     source          1..212
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /mol_type="genomic DNA"
                     /clone="[illegible]"
                     /clone_lib="OSU1 Rice BAC Library (Boo01)"
                     /tissue_type="leaf"
                     /lab_host="E. coli DH10B"
                     /db_xref="taxon:4530"
                     /note="Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Molecular studies in rice is a logical part (or 24) with a
                     haploid genome equivalent of 45[illegible] Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times smaller than that of Arabidopsis, makes it suitable
                     for genetic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using HindIII as the cloning enzyme. The
                     library contains [illegible] clones with an average insert size
                     of 150 Kbp yielding approximately 9 haploid genome
                     equivalents. The deep coverage allows the isolation of a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (3 dry spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.rice.genome.clemson.edu)."
```

```
BASE COUNT  [illegible]
ORIGIN


//

LOCUS       [illegible]
DEFINITION  [illegible]
ACCESSION   [illegible]
VERSION     [illegible]
KEYWORDS    HTG.
SOURCE      [illegible]
  ORGANISM  [illegible]
REFERENCE   [illegible]
  AUTHORS   [illegible]
  TITLE     [illegible]
  JOURNAL   [illegible]
COMMENT     [illegible]
FEATURES             Location/Qualifiers
     source          [illegible]
BASE COUNT  [illegible]
ORIGIN
```



```
                     Further details
                     of mouse tissues.
FEATURES             Location/Qualifiers
     source          1..642
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5003I18EI2"
                     /clone_lib="RIKEN full-length enriched, 11 days pregnant
                     adult female ovary and uterus"
                     /sex="female"
                     /tissue_type="ovary and uterus"
                     /dev_stage="11 days pregnant adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAATAGAAAAGTCGTTTTTTTTTTTTTTTVN 3']; cDNA was
                     prepared by using trehalose thermoactivated reverse
                     transcriptase and subsequently enriched for full length by
                     cap trapper. cDNA went through one round of normalization
                     to Rot 1.0 and subtraction to Rot 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     5'-[illegible]-3'; cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(-) after bulk excision
                     from Lambda FLC I. *Cloning sites, 5' end: SalI; 3' end:
                     BamHI."
BASE COUNT       192 a    140 c    177 g    133 t
ORIGIN


Alignment Scores:
            Length:  523.00              Identity:      60
             Ratio:  2.562                 Gaps:        0
 Percent Similarity:  80.688   Percent Identity:  66.667

Alignment block:
 US-09-817-762-2-SEQ-ID-76 x BP037489  ..

 Align Set 1/1    Seq: BP037489  From: 1  To: 642

       1 MetLysLysLysSerLeuSerLeuProPheProLysLeuPheSerSerAsA  33
         ||||||||| |||  ||||||     |||  |||||  ||||||||||||
     209 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  260

      34 aAspLysPheAspTyrLeuLeuMetPheValGlySerLeuGlyAlaIleV  80
         |||  |||||  |||  ||||||||||||||||||||||||||||||||||
     261 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  310

      81 aHisLysSerSerMetProValPheIleLeuIleSerTyrMetVal  125
         |||||||||||||||||||||||||||||||||||||||||||||||
     311 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  360

     126 AsnGlyPheGlyLysAsnLeuMetAspLeu  154
         |||||||||||||||||||||||||||||||
     361 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAA  390

Sequence: db_est2:BP037489

Sequence Annotation Block:
LOCUS       BP037489                 523 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  BVSME1003A16f Hordeum vulgare spike EST Library BVSNA0012
            (Fusarium infected) Hordeum vulgare cDNA clone BVSME1003A16f, mRNA
            sequence.
ACCESSION   BP037489
VERSION     BP037489.1  GI:18208004
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Wise, M., Muehlbauer, G., Kleinhofs, A., Ullrich, S.,
            Graner, A., Baum, M. and Matthews, D.
  TITLE     Expressed Sequence Tags (ESTs) generated from Fusarium
            infected barley spikes
  JOURNAL   Unpublished
COMMENT     Contact: Wise, M.
            Clone: 5003I18EI2
            Clone lib: RIKEN full-length enriched, 11 days pregnant adult
            female ovary and uterus
            cDNA: Trehalose thermoactivated reverse transcriptase and
            subsequently enriched for full length by cap trapper
            Seq primer: GAGAGAGAAATAGAAAAGTCGTTTTTTTTTTTTTTTVN
            Host: DH10B
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /db_xref="taxon:4513"
                     /clone="BVSME1003A16f"
                     /clone_lib="BVSNA0012"
                     /dev_stage="spike"
                     /tissue_type="spike"
                     /note="Vector: pBluescript
                     KS(-); Insert was BE at
                     Morphological data: spike
                     at the early to mid
                     uninucleate stage;
                     Developmental stage of
                     infected tissue: Early
                     cDNA pools were size
                     fractionated before
                     ligation into the pBS(-)
                     KS( ) cDNA plasmid
                     vector. Clones were picked
                     randomly. Sequences were
                     edited and assembled with
                     Sequencher 3.0. Kinds and
                     Wise, Yu, Ellis, Beaty,
                     Caldwell, Bayles, Chen,
                     Kleinhofs, Graner, Baum,
                     Ullrich, Matthews, Muehlbauer
                     and Wise, unpublished."
BASE COUNT       192 a    140 c    177 g    133 t
ORIGIN


Alignment Scores:
            Length:  523.00              Identity:      60
             Ratio:  2.562                 Gaps:        0
 Percent Similarity:  80.688   Percent Identity:  66.667

Alignment block:
 US-09-817-762-2-SEQ-ID-76 x BP037489  ..

 Align Set 1/1    Seq: BP037489  From: 1  To: 642

       1 MetLysLysLysSerLeuSerLeuProPheProLysLeuPheSerSerAsA  33
         ||||||||| |||  ||||||     |||  |||||  ||||||||||||
     209 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  260

      34 aAspLysPheAspTyrLeuLeuMetPheValGlySerLeuGlyAlaIleV  80
         |||  |||||  |||  ||||||||||||||||||||||||||||||||||
     261 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  310

      81 aHisLysSerSerMetProValPheIleLeuIleSerTyrMetVal  125
         |||||||||||||||||||||||||||||||||||||||||||||||
     311 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  360

     126 AsnGlyPheGlyLysAsnLeuMetAspLeu  154
         |||||||||||||||||||||||||||||||
     361 ATTTTTTTTAAAAAAAAAAAAAAAAAAAAAA  390

Sequence: db_est2:BP037489

Sequence Annotation Block:
LOCUS       BP037489                 523 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  BVSME1003A16f Hordeum vulgare spike EST Library BVSNA0012
            (Fusarium infected) Hordeum vulgare cDNA clone BVSME1003A16f, mRNA
            sequence.
ACCESSION   BP037489
VERSION     BP037489.1  GI:18208004
KEYWORDS    EST.
SOURCE      barley.
```

```
                  ...ellular tissues.
  FEATURES             Location/Qualifiers
       source          1..284
                       /db_xref="taxon:10090"
                       /clone="RIKEN full-length enriched, adult male liver
                       tumor"
                       /sex="male"
                       /tissue_type="liver tumor"
                       /dev_stage="adult"
                       /lab_host="DH10B"
                       /note="Sites 2: SacII Site at BamHI cDNA library was
                       prepared and sequenced in Mouse Genome Encyclopedia
                       Project of Genome Exploration Research Group in Riken
                       Genomic Sciences Center and Genome Science Laboratory in
                       RIKEN, Division of Experimental Animal Research in Riken
                       contributed to prepare mouse tissues. 1st strand cDNA was
                       primed with a primer [5'
                       GAATTCGAGCTCGGTACCAAGCTTGATCTTTTTTTTTTTTTTTVN 3'], cDNA was
                       prepared by using trehalose thermoactivated reverse
                       transcriptase and subsequently enriched for full-length by
                       cap trapper. Second strand cDNA was prepared with the
                       primer adapter of sequence [5'
                       TAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGNNNNNNNNN 3']. cDNA
                       was excised with BamHI and XhoI. Vector: a modified
                       pBluescript KS(+) after bulk excision from Lambda ZAP II.
                       Tissue was provided by William A. Boyd, Roswell Park
                       Cancer Institute, Department of Molecular and Cellular
                       Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose
                       assistance we gratefully acknowledge."
BASE COUNT        [illegible]  a  [illegible]  c  [illegible]  g  [illegible]  t  [illegible]  others
ORIGIN
```

```
Alignment scores:
            Quality:  117.0                 Length:     56
              Ratio:  2.059                   Gaps:      0
 Percent Similarity:  98.571   Percent Identity: 41.071

Alignment block:
US-09-817-762-2_COPY_1_76 X BB057985  1..284

 Align Seq 1/2  1 to BB057985  Frame: 1  to 284

      17 GlyLysLysLysLysLeuSerLeuArgTheThrLysLeuTheSerProAl 53
         ||||||||||||  ||| ||| ||||||  ||||||  |||||
      273 AAGAAGAAGAAAGTCAATTAAGTTTAAGAACTCCGGATACTC 230

      54 AspLysPheAspTyrLeuLeuMetPheValGlySerLeuGlyAlaIleV 99
         |||||  |||  ||||||||||||||||||||  ||||||
      284 TGATTTTGATTATTTAATGTTTGTTGGTTCTCTAGGTGCTATC 327

      100 GlyLeuSerSerMetProValPheIleLeuPheGlyGlnMetVal 146
          |||||||||||  ||||||  ||||||||||||||||||  ||||||||||
      328 TTCATTGGATCAATGCCTGTTTTTATTCTTTTTGGACAAATGGT 373

      147 AsnGlyPheLysLysAsn 162
          |||  |||  |||  |||
      374 GATAATTTTAAGAAAAT 391

No alignments above BASELINE SCORE

No abbreviation block:
LOCUS       CNS00XCF                 393 bp    DNA     linear   GSS 29-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence T7 end of BAC T14P5 of
            IAMU library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL049743
VERSION     AL049743.1  GI:5834508
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
```

```
            Eukaryota; [illegible]; Viridiplantae; Streptophyta;
            Spermatophyta; Magnoliophyta; [illegible]
            Posted date: [illegible] 29-JUN-1999
REFERENCE   1  (bases [illegible])
  AUTHORS   Salanoubat,M., Lemcke,K., Rieger,M., Ansorge,W., Unseld,M.,
            Fartmann,B., Valle,G., Blocker,H., Perez-Alonso,M. and Obermaier,B.
  JOURNAL   Unpublished
REFERENCE   2  (bases [illegible])
  AUTHORS   Salanoubat,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1999) Genoscope, BP
            191, 91006 Evry Cedex, FRANCE (E-mail: [illegible]@genoscope.cns.fr,
            Tel: [illegible], Fax: [illegible])
FEATURES             Location/Qualifiers
     source          1..393
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="T14P5"
                     /clone_lib="IAMP"
                     /note="Sites 2: SacII Site at BamHI cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
BASE COUNT        [illegible]  a  [illegible]  c  [illegible]  g  [illegible]  t  [illegible]  others
ORIGIN
```

```
Alignment scores:
            Quality:  [illegible]                 Length:     [illegible]
              Ratio:  [illegible]                   Gaps:      [illegible]
 Percent Similarity:  [illegible]   Percent Identity: [illegible]

Alignment block:
US-09-817-762-2_COPY_1_76 X [illegible]  1..[illegible]

 Align Seq 1/2  1 to [illegible]  Frame: [illegible]  to [illegible]

      [illegible]

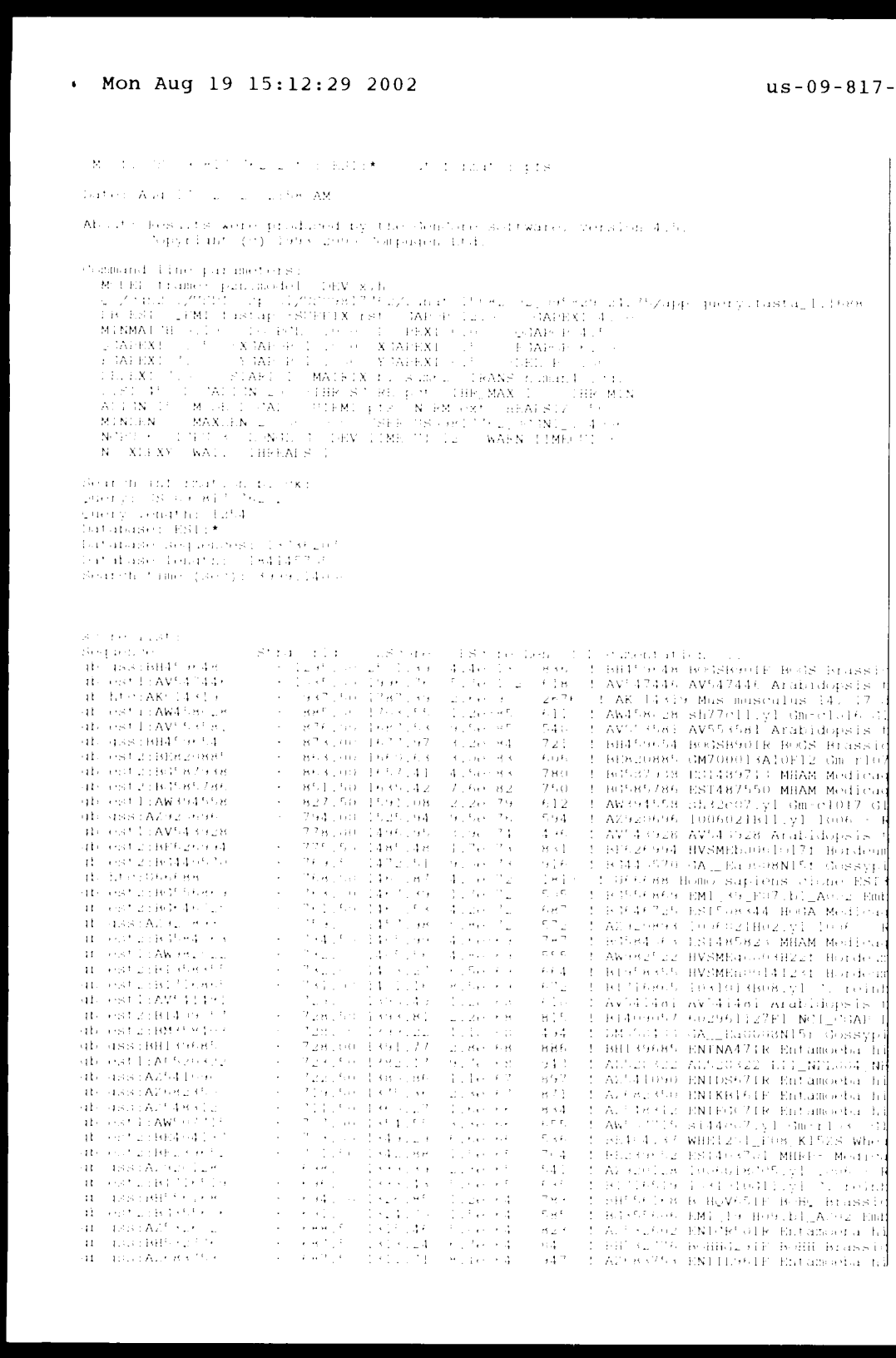
      [illegible]

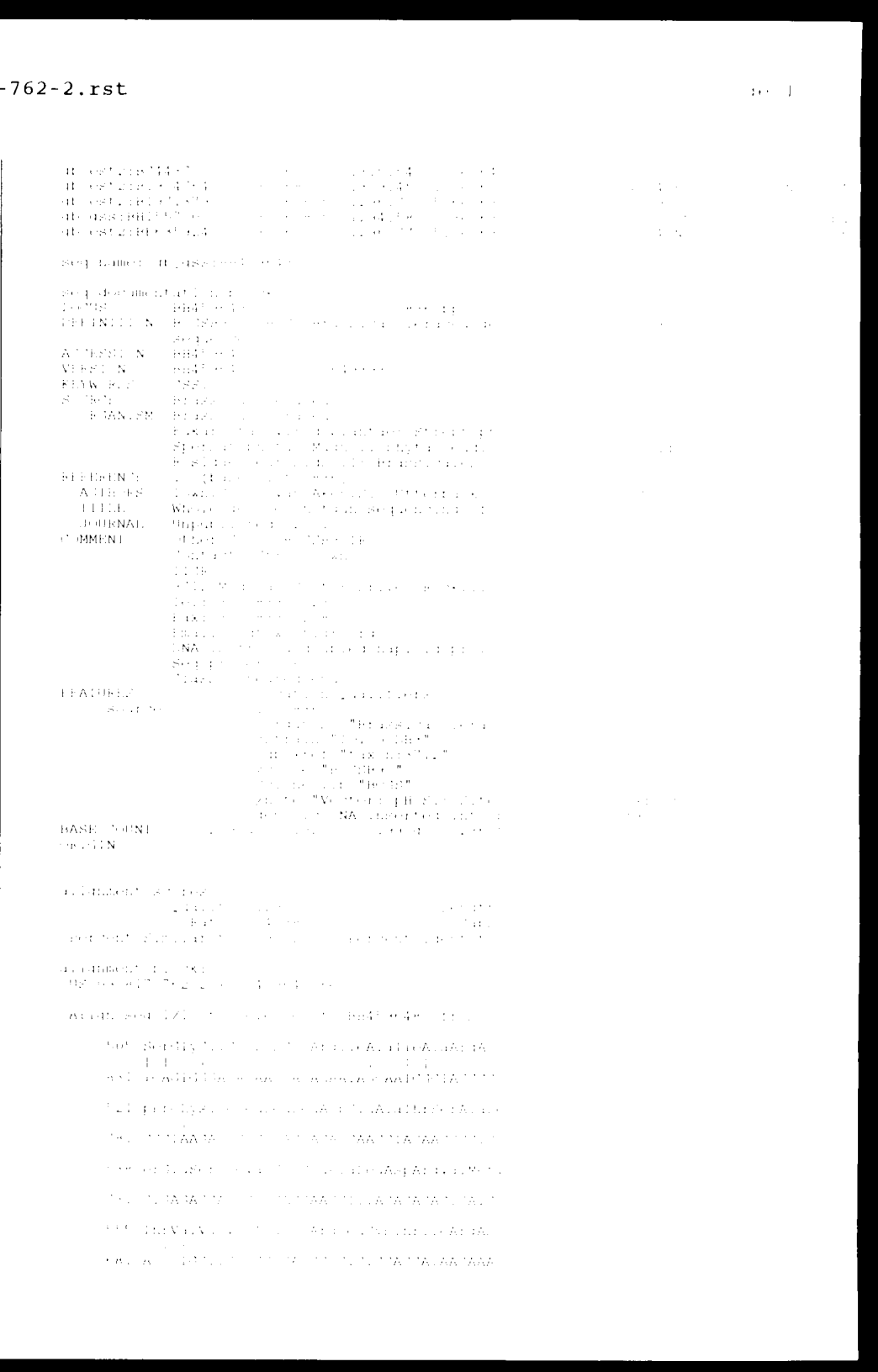
      [illegible]

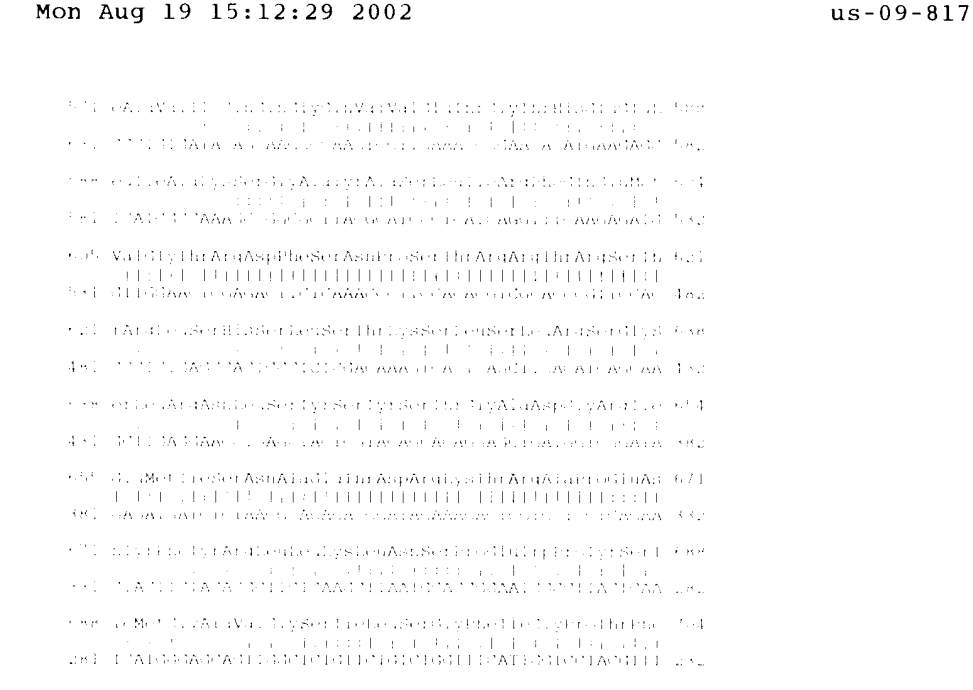
      [illegible]

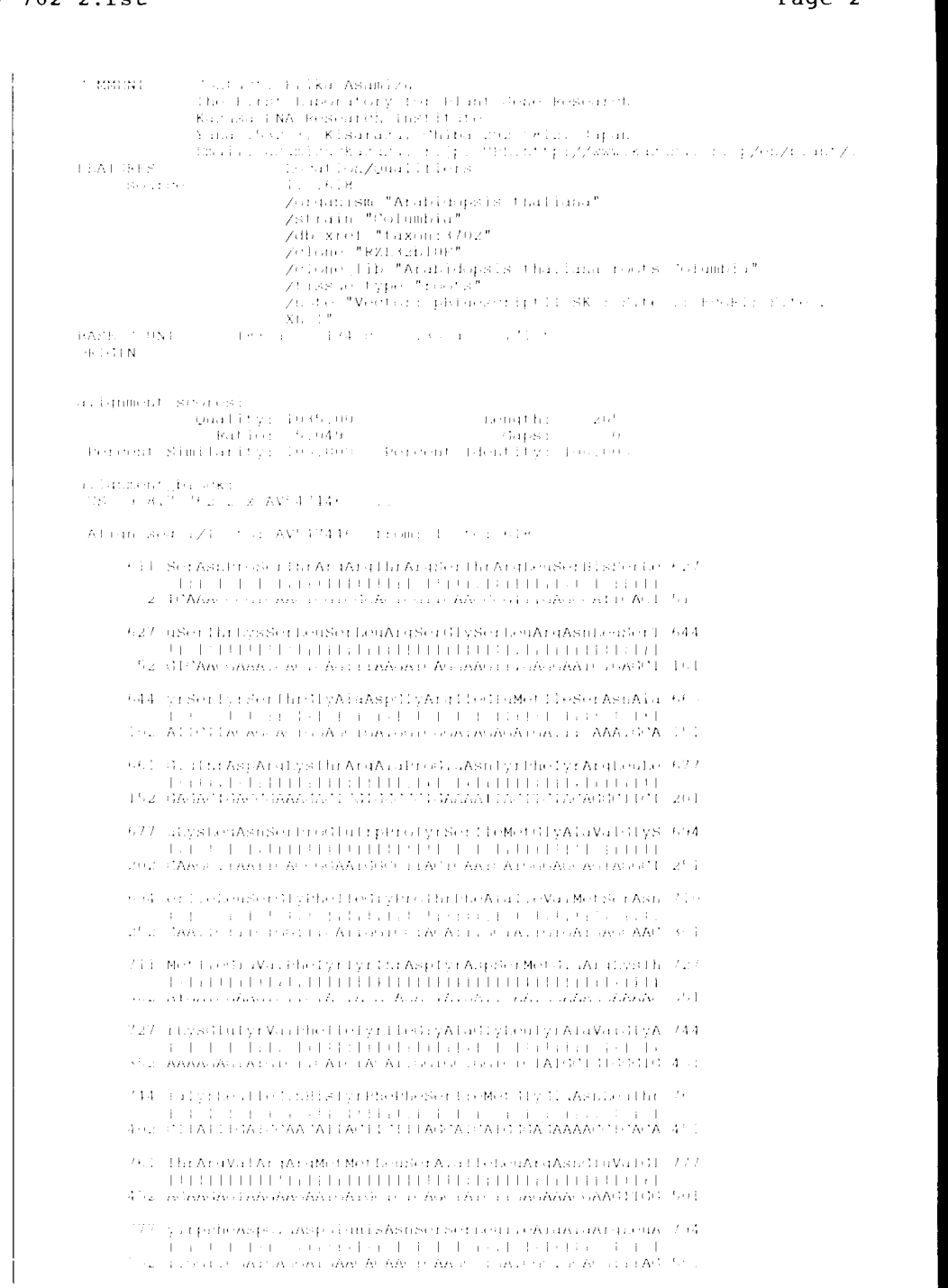
No alignments above BASELINE SCORE

No abbreviation block:
```









```
794 IleThrAspAlaAlaAspValLysSerAlaIleAlaGlnArgIleSerVal 810
552 [illegible]AGA 601

811 IleLeuGlnAsnMet 815
602 ATT[illegible]CAAAA[illegible]ATG 616

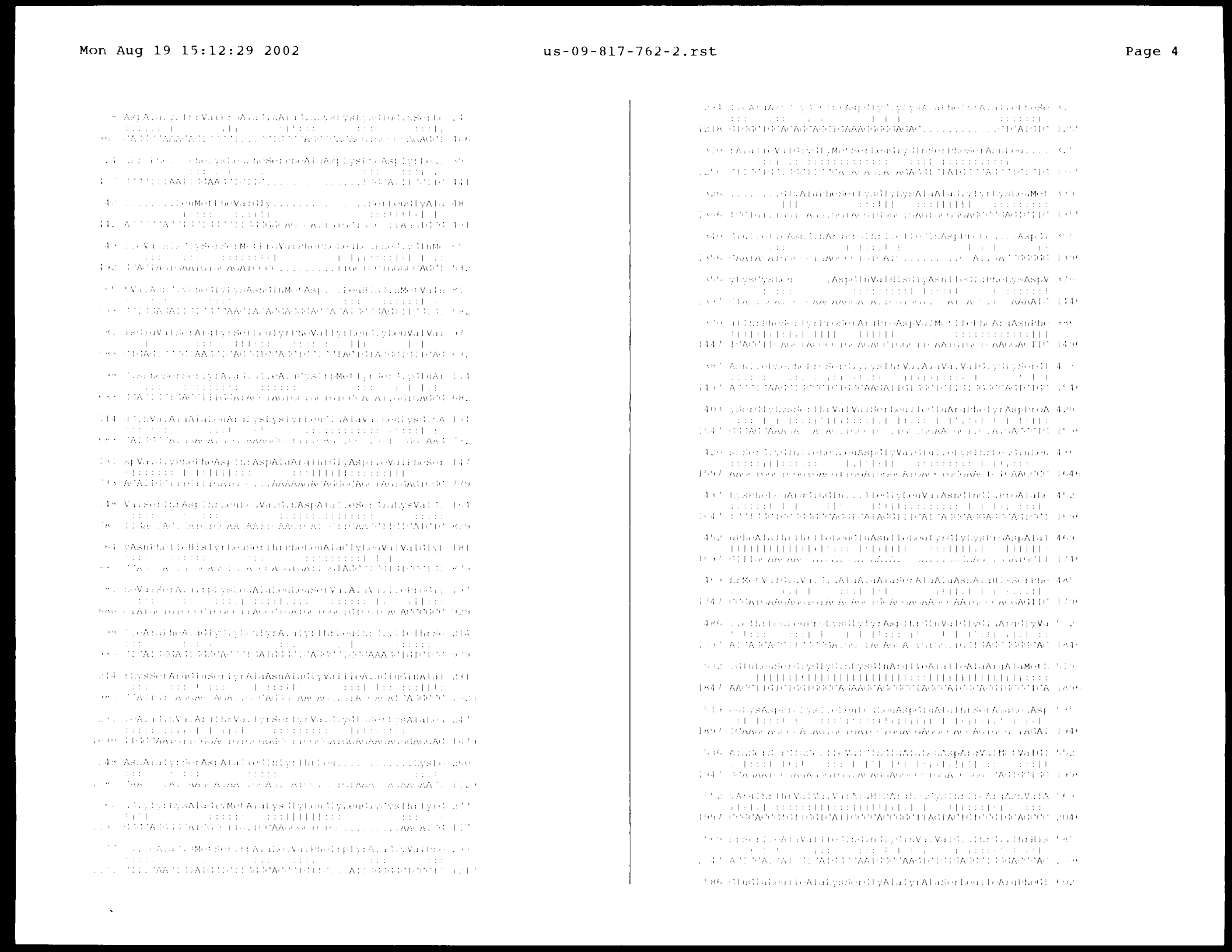
Seq name: gb_htc:AK014319

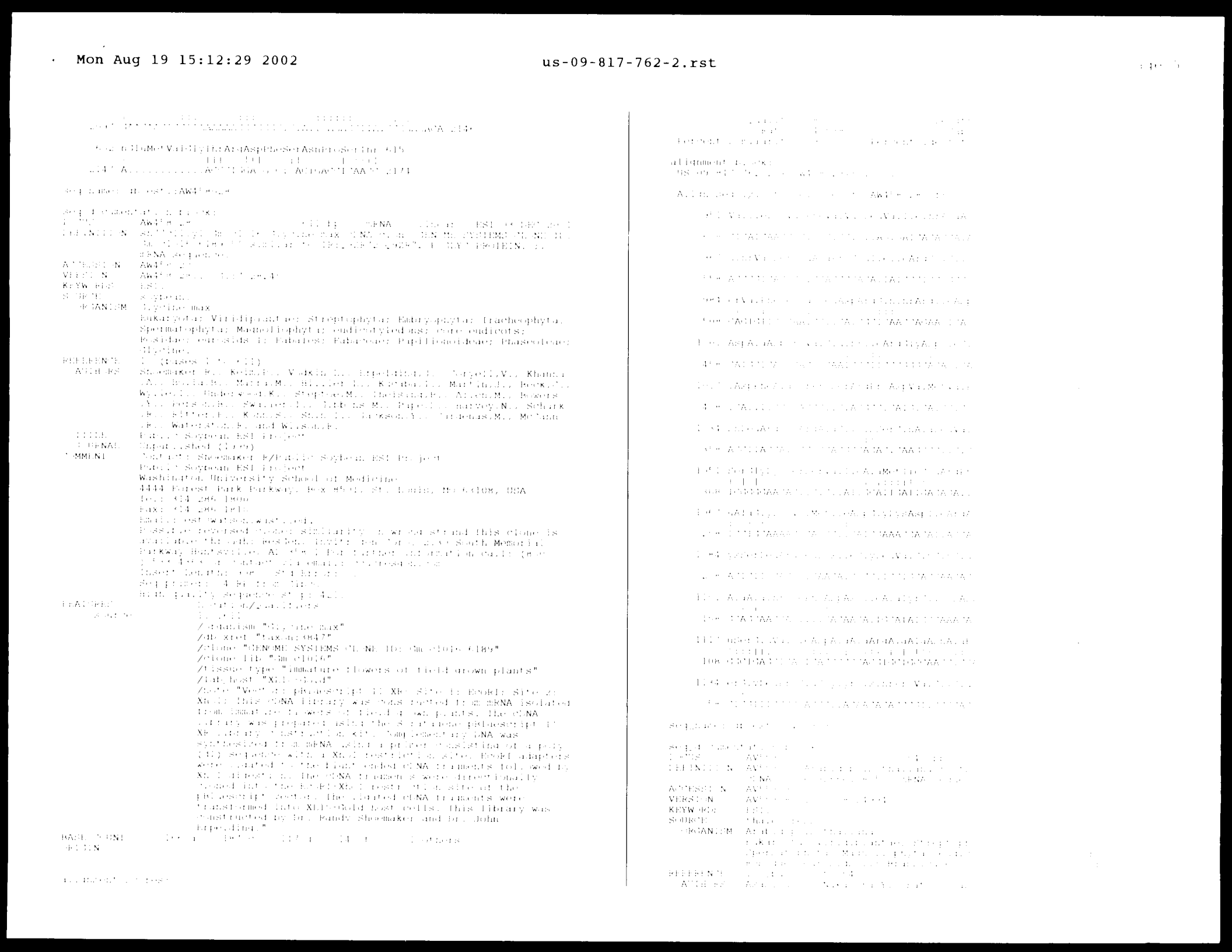
Seq documentation block.
LOCUS       AK014319                2676 bp    mRNA    linear   HTC 19-JAN 2002
DEFINITION  Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
            enriched library, clone:3222401P09:ATP-binding cassette, sub-family
            B (MDR/TAP), member 8, full insert sequence.
ACCESSION   AK014319
VERSION     AK014319.1  GI:12852009
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 14, 17 days embryo head cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:3222401P09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED    11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2676)
  AUTHORS   [illegible], Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kanukawa,I.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), [illegible]
            Exploration Research Group, RIKEN [illegible]
            RIKEN [illegible] Institute [illegible]
            Kanagawa [illegible] Japan ([illegible])
            URL:http://[illegible]riken.[illegible]
            Fax:[illegible])
COMMENT     Please [illegible] web site (http://[illegible]
            further details.
            cDNA [illegible] was prepared and sequenced [illegible]
            Embryo [illegible] Genome Exp[illegible]
            [illegible] Center and [illegible]
            [illegible] Experimental Animal [illegible]
            prepare[illegible] First strand [illegible]
            [illegible]AGAGAGAGAA[illegible]AAGAG[illegible]
            prepared [illegible] trehalose [illegible]
            and [illegible] enriched for [illegible]
            strand cDNA was prepared with [illegible]
            GAGA[illegible]AATAAAT[illegible]
            with [illegible] XhoI [illegible]
            selected [illegible] Vector [illegible]
            [illegible] lambda [illegible]
            [illegible]
FEATURES             Location/Qualifiers
     source          [illegible]
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:[illegible]"
                     /db_xref="taxon:[illegible]"
                     /clone="3222401P09"
                     /tissue_type="head"
                     /clone_lib="RIKEN [illegible]"
                     /dev_stage="14, 17 [illegible]"
     gene            [illegible]
                     /gene="Abcb8"
     CDS             [illegible]
                     /gene="Abcb8"
                     /note="ATP-binding cassette [illegible]
                     [illegible]MGD[illegible]
                     [illegible]"
                     /codon_start=1
                     /protein_id="BAB29[illegible]"
                     /db_xref="GI:1285[illegible]"
                     /translation="MLVHLFR[illegible]
                     [illegible]VAL[illegible]FAHLPRAFFASHWSFSA[illegible]
                     [illegible]FENWKLFWHLLRHLL[illegible]
                     [illegible]GHMGSFVSESRKLSVGLLLYGV[illegible]
                     [illegible]
                     [illegible]ALMGVGILMGS[illegible]
                     [illegible]
                     [illegible]IVQRSMASLSVLE[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]"
     polyA_signal    [illegible]
                     /note="[illegible]"
     polyA_site      [illegible]
                     /note="[illegible]"
BASE COUNT   [illegible]
ORIGIN

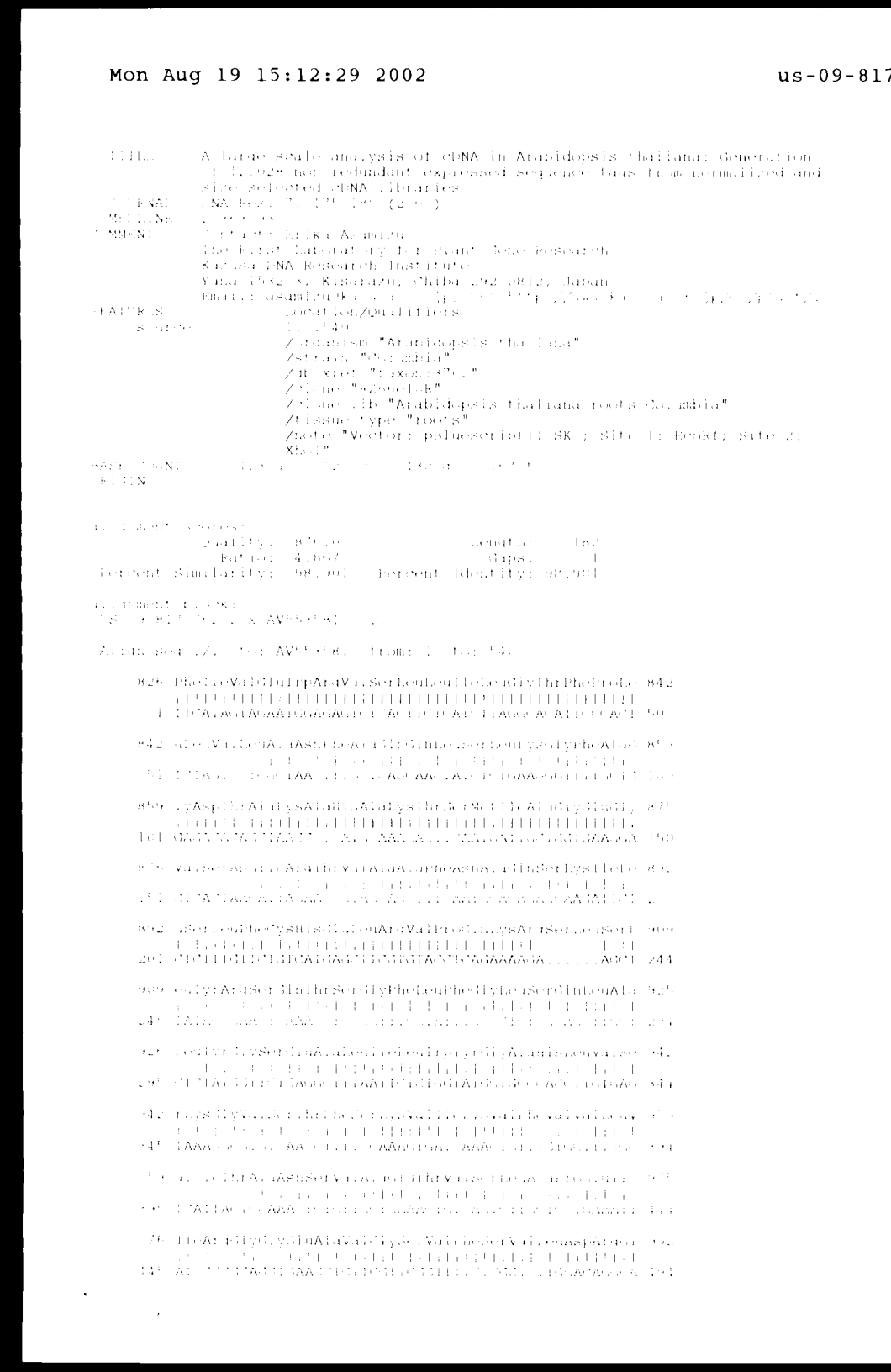
Alignment Scores:
[illegible]
Percent Similarity: [illegible]

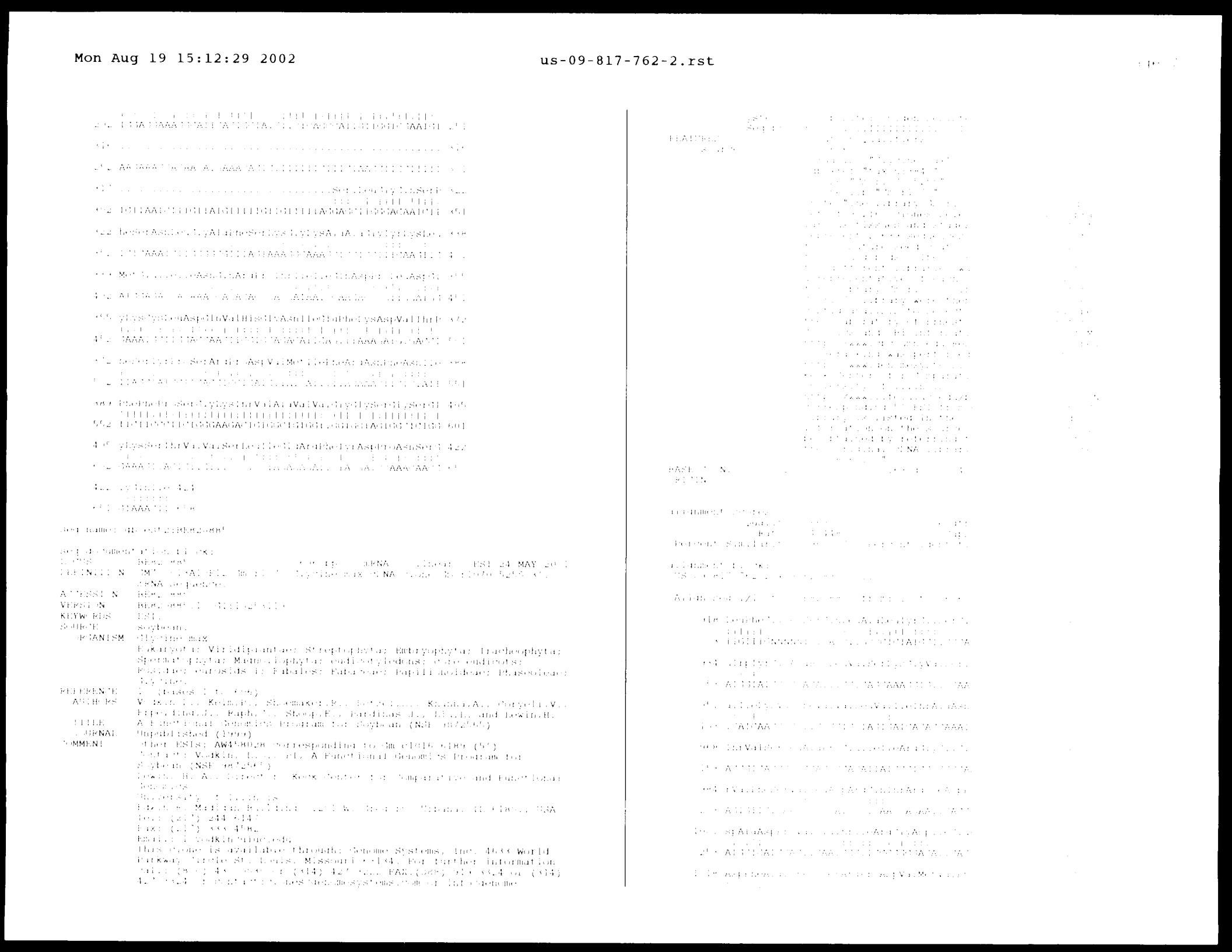
Alignment Block:
US-09-817-762-2 [illegible]

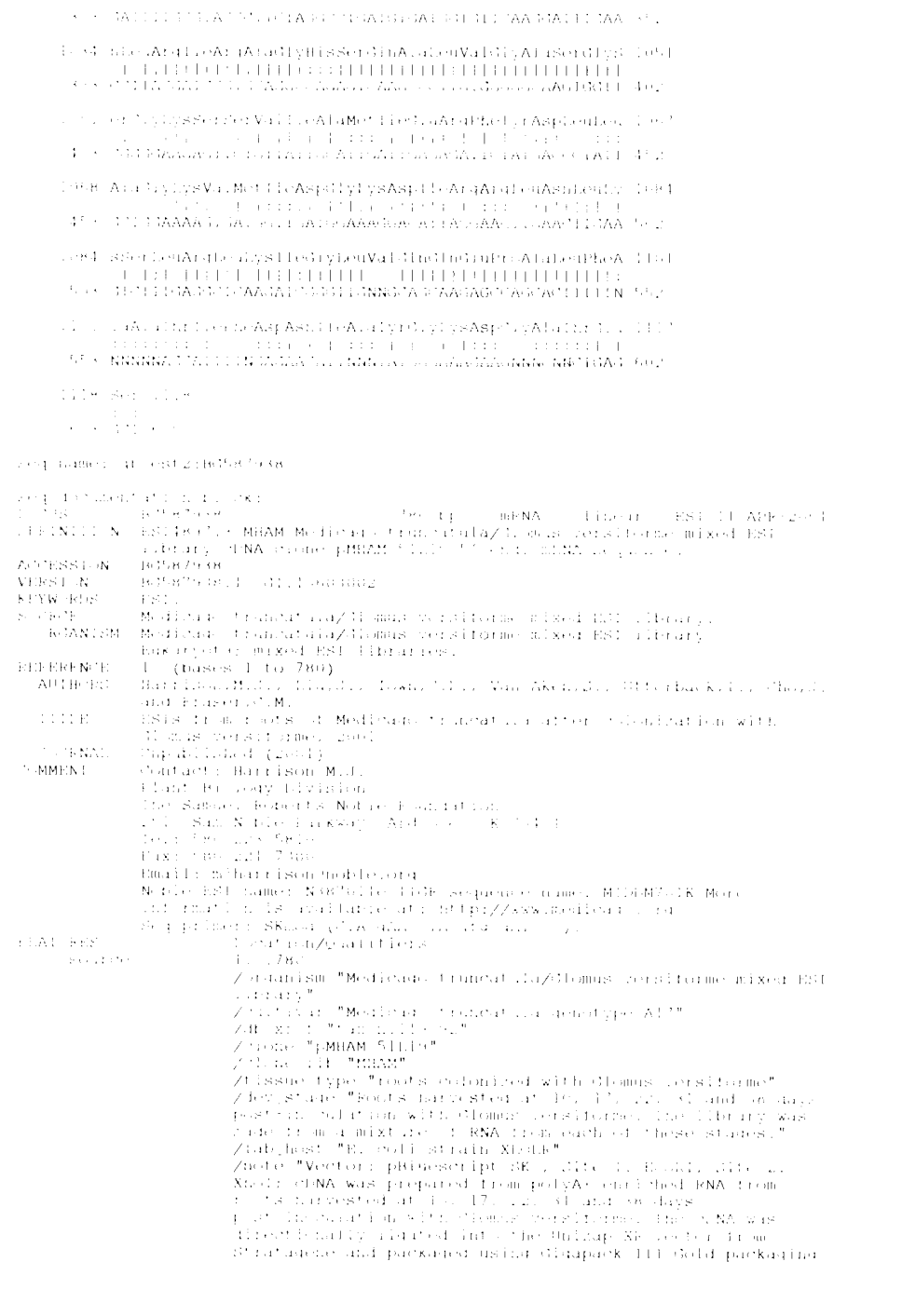
Align seq 1/1 [illegible]
```



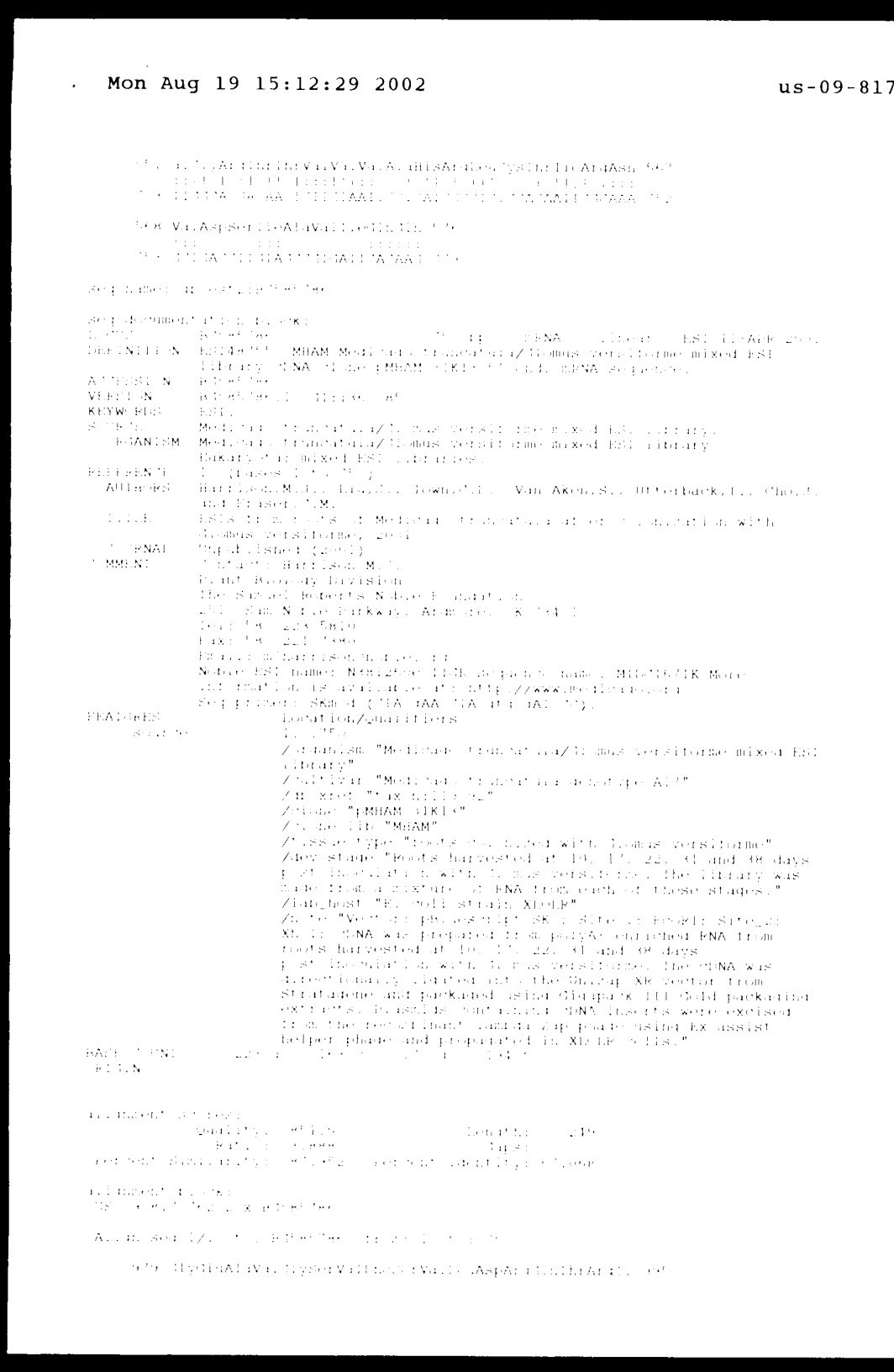








BASE COUNT 244 a 156 c 184 g 206 t
ORIGIN
Alignment Scores:
Quality: 893.00 Length: 899
Ratio: 3.704 Gaps: 0
Percent Similarity: 89.961 Percent Identity: 66.963
Alignment Blocks:
Align Seq 1/1 to: BG587938 from: 1 to: 780
DEFINITION MHAM Medicago truncatula/Glomus versiformis mixed EST
library, mRNA sequence.
ACCESSION BG587938
VERSION BG587938.1 GI:13668802
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiformis mixed EST library.
ORGANISM Medicago truncatula/Glomus versiformis mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 780)
AUTHORS Harrison,M.J., Dixon,R.A., Van Buuren,M., Dickstein,R., Bell,C.,
and Fraser,S.M.
TITLE EST's from roots of Medicago truncatula after colonization with
Glomus versiformis
JOURNAL Unpublished (2002)
COMMENT Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ard... OK 7...
Tel: 580-...-...
Fax: 580-221-7380
Email: mharrison@noble.org
Notes: EST names: MJH...f10b, sequence names: MJH...
...ral mRNA's is available at: http://www.noble...
Seq primer: SK...
FEATURES Location/Qualifiers
source 1..780
/organism="Medicago truncatula/Glomus versiformis mixed EST
...ibrary"
/clone="Medicago truncatula cDNA prototype A1"
/db_xref="taxon:...."
/clone="pMHAM S1L1r"
/clone_lib "mham"
/tissue_type "roots colonized with Glomus versiformis"
/dev_stage "Roots harvested at 10, 17, 24, 31 and 38 day...
post-in... colonization with Glomus versiformis. The library was
made from a mixture of 5 RNA from each of these stages."
/lab_host "E. coli strain XL1-B"
/note "Vector: pBluescript SK-; Cla... I; BamHI, Cla... ...
XbaI; cDNA was prepared from polyA+ enriched RNA from
... 's harvested at 10, 17, ... 31 and 38 days
after inoculation with Glomus versiformis. The cDNA was
directionally ligated into the BamHI-XhoI sites in ...
Propagated and packaged using Gigapack III Gold packaging
extract. Plasmids were transformed into cDNA insert was sequenced
from the recombinant lambda Zap phagemid in XL1-B
helper phage and propagated in XL1-B cells."



Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 612)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Retzel,E., Erpelding,J., Coryell,V., Khanna,A., Bozzola,R., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Research Genetics Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call (800) 533-4363 or contact via email: info@resgen.com
Insert Length: 1341 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1..612
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-3421"
/clone_lib="Gm-c1017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10 Gold"
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10 Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."
BASE COUNT 149 a 118 c 187 g 158 t
ORIGIN

Alignment Scores:
Quality: 827.50 Length: 204
Ratio: 4.222 Gaps: 2
Percent Similarity: 96.078 Percent Identity: 93.824

Alignment Seq:
US-09-817-762-2 x AW394558 ...

Alignment [illegible] AW394558 [illegible]

[illegible]

1023 [illegible]AspValMetValPheArgAsp[illegible]ArgAla[illegible] 1040
[illegible] 256

[illegible]AlaSerGlySerGly[illegible]SerVal [illegible]
[illegible]AAG[illegible]GCGG[illegible]AAGAG[illegible] 306

1057 IleAlaMetIle[illegible]ArgPhe[illegible]AspLeu[illegible]LysValMet[illegible] 1073
[illegible]AAAAG[illegible] 356

[illegible]AspGlyTyrAspIleArgArg[illegible]Asn[illegible]LysSer[illegible] [illegible]
[illegible] 406

1090 [illegible]ProAla[illegible]AlaThrIlePheAsp [illegible]
[illegible] 456

1107 Asn[illegible]GlyLysAsp[illegible]AlaThrGluSer[illegible]ValIleAspAl [illegible]
[illegible]GATTGAGGC 506

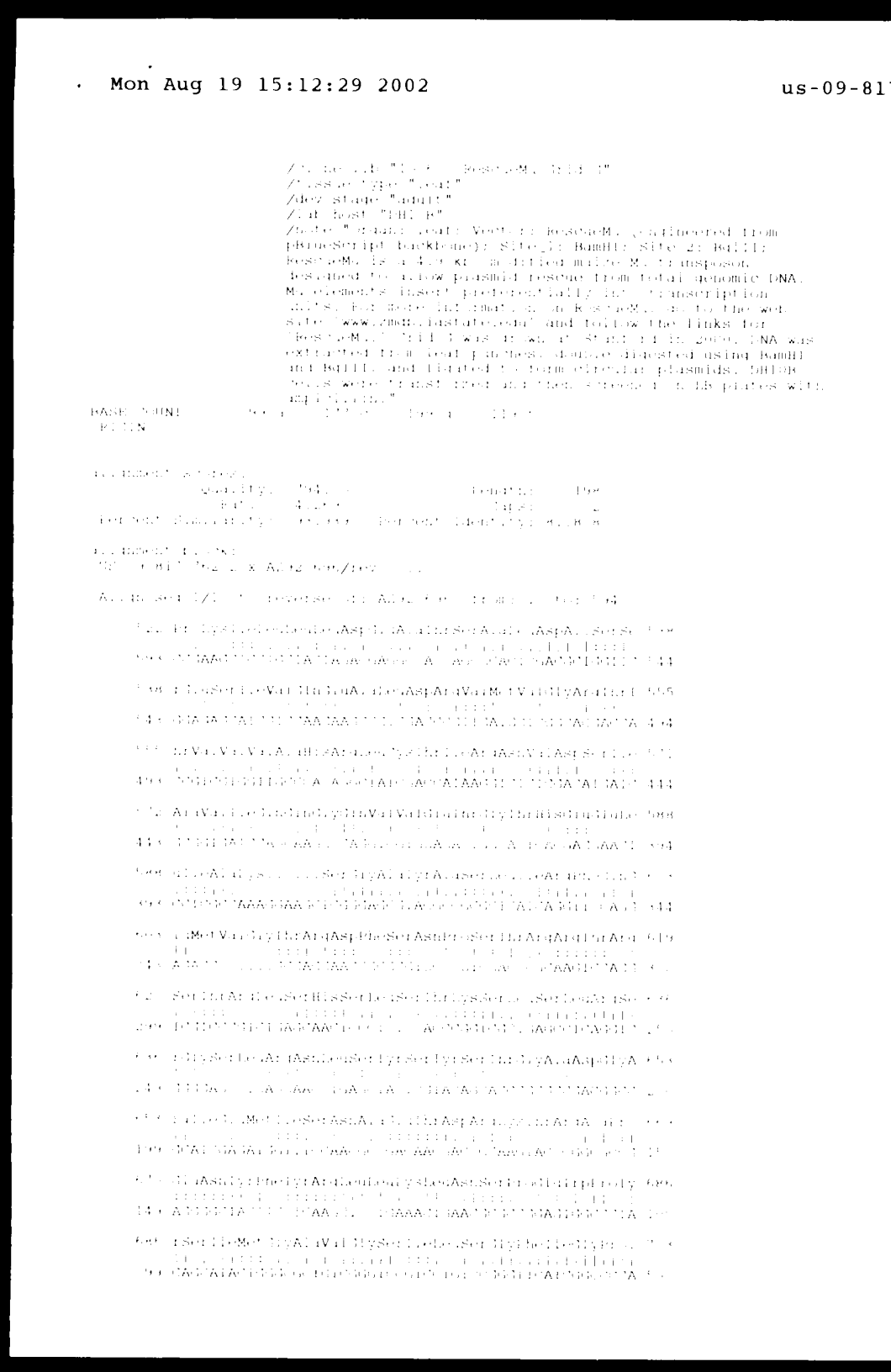
1123 [illegible]AlaAsnAlaHisGlyPheIleSerGlyLeu[illegible]Gly 1139
[illegible] 556

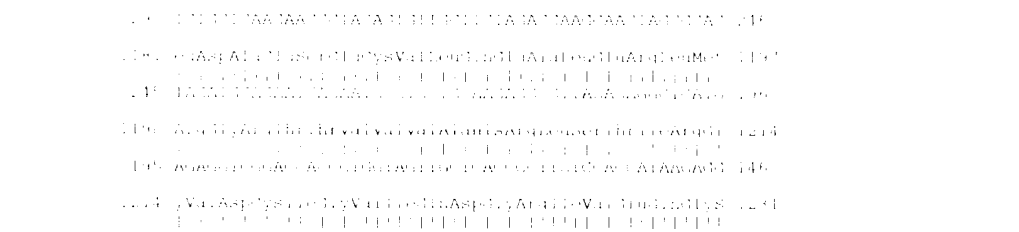
1140 [illegible]ProVal[illegible]ArgGlyValGln[illegible]SerGlyGly[illegible] 1155
[illegible]AAAG 602

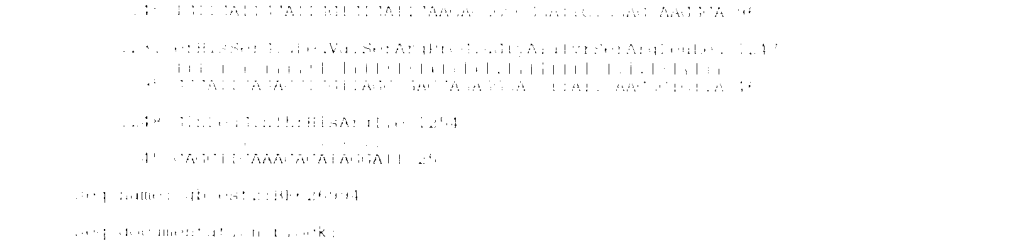
1156 sGlnArgIle 1159
[illegible]AAAAGATT[illegible] 612

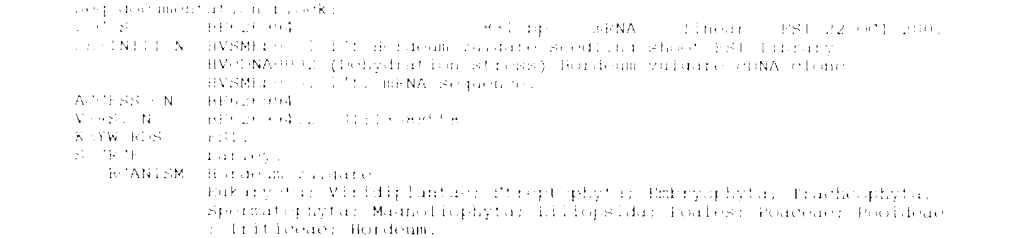
Seq name: gb_gss:AZ920696

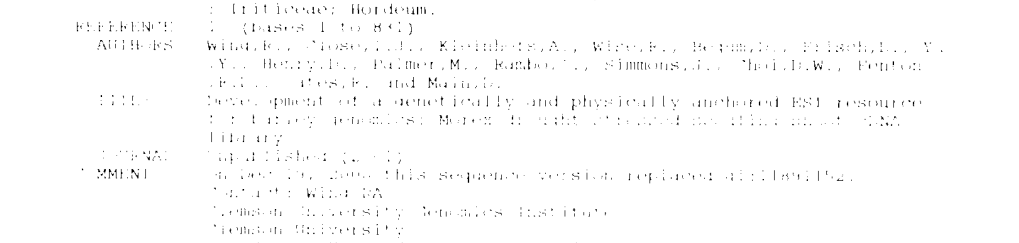
seq_documentation_block:
LOCUS AZ920696 594 bp DNA linear GSS [illegible]
DEFINITION [illegible] RescueMu Grid C Zea mays genomic DNA sequence.
ACCESSION AZ920696
VERSION AZ920696.1 GI:13397594
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 594)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
385 Serra Mall [illegible] Ave, Palo Alto, CA 94305, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases. Reverse complemented post ligation sequence from source sequence.
Plate: [illegible] row: 38
Class: transposon-tagged.
FEATURES Location/Qualifiers
source 1..594
/organism="Zea mays"
/cultivar="mixed background W22/A188/B73"
/db_xref="taxon:4577"

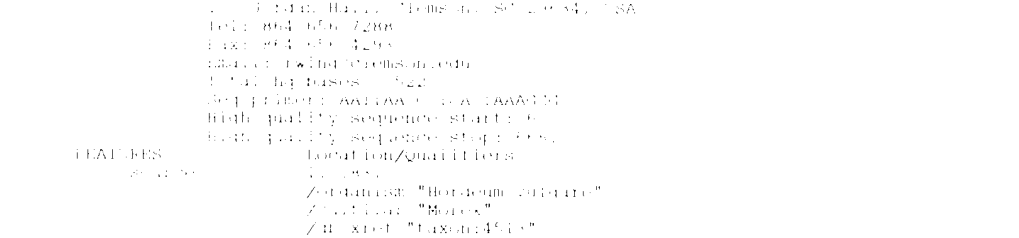


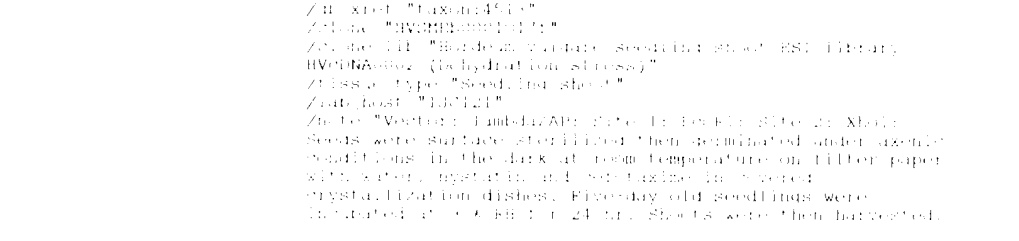




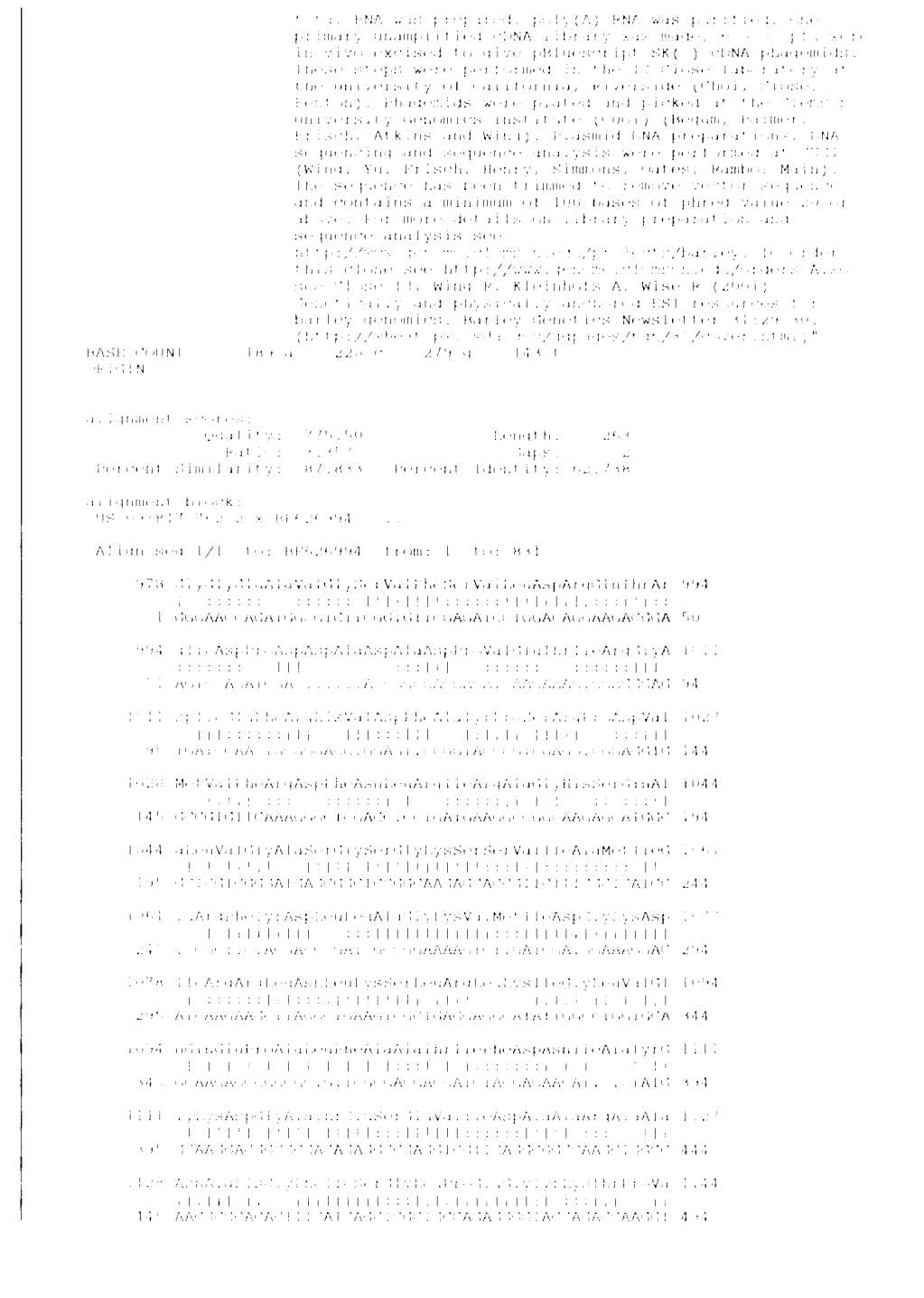


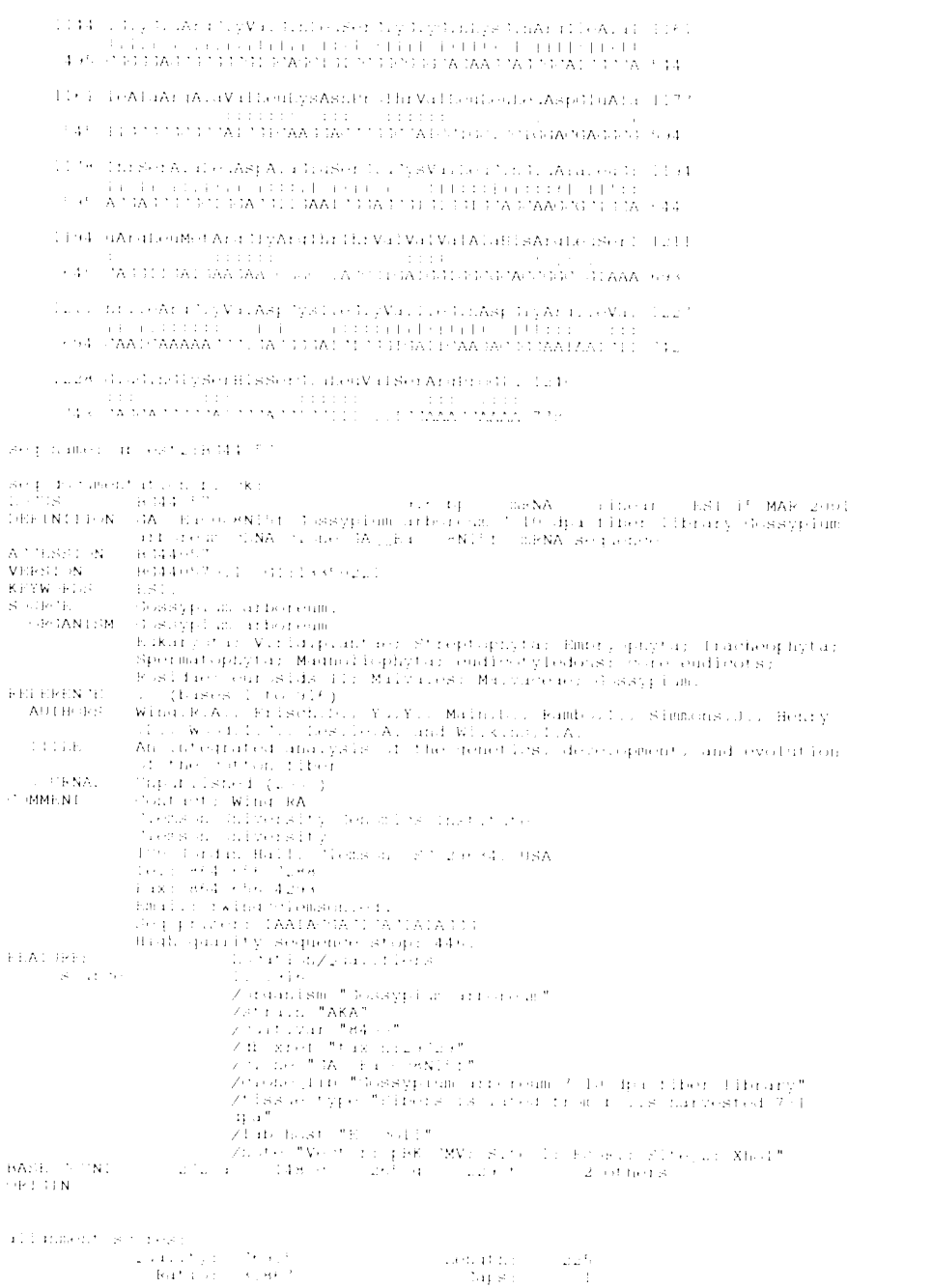


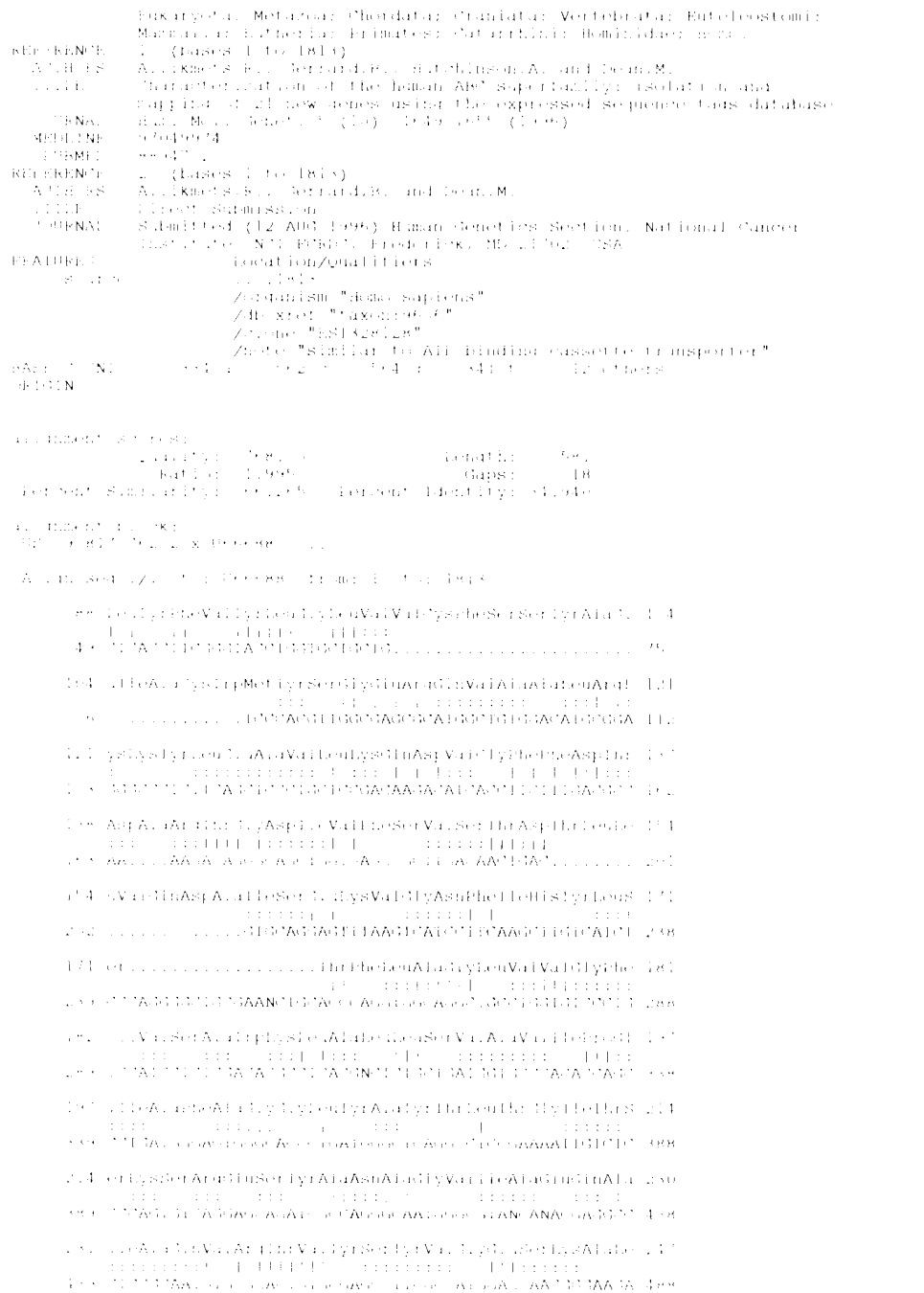


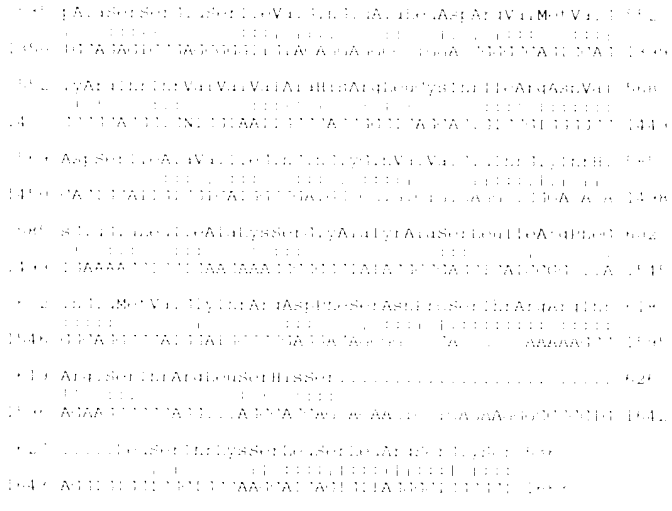


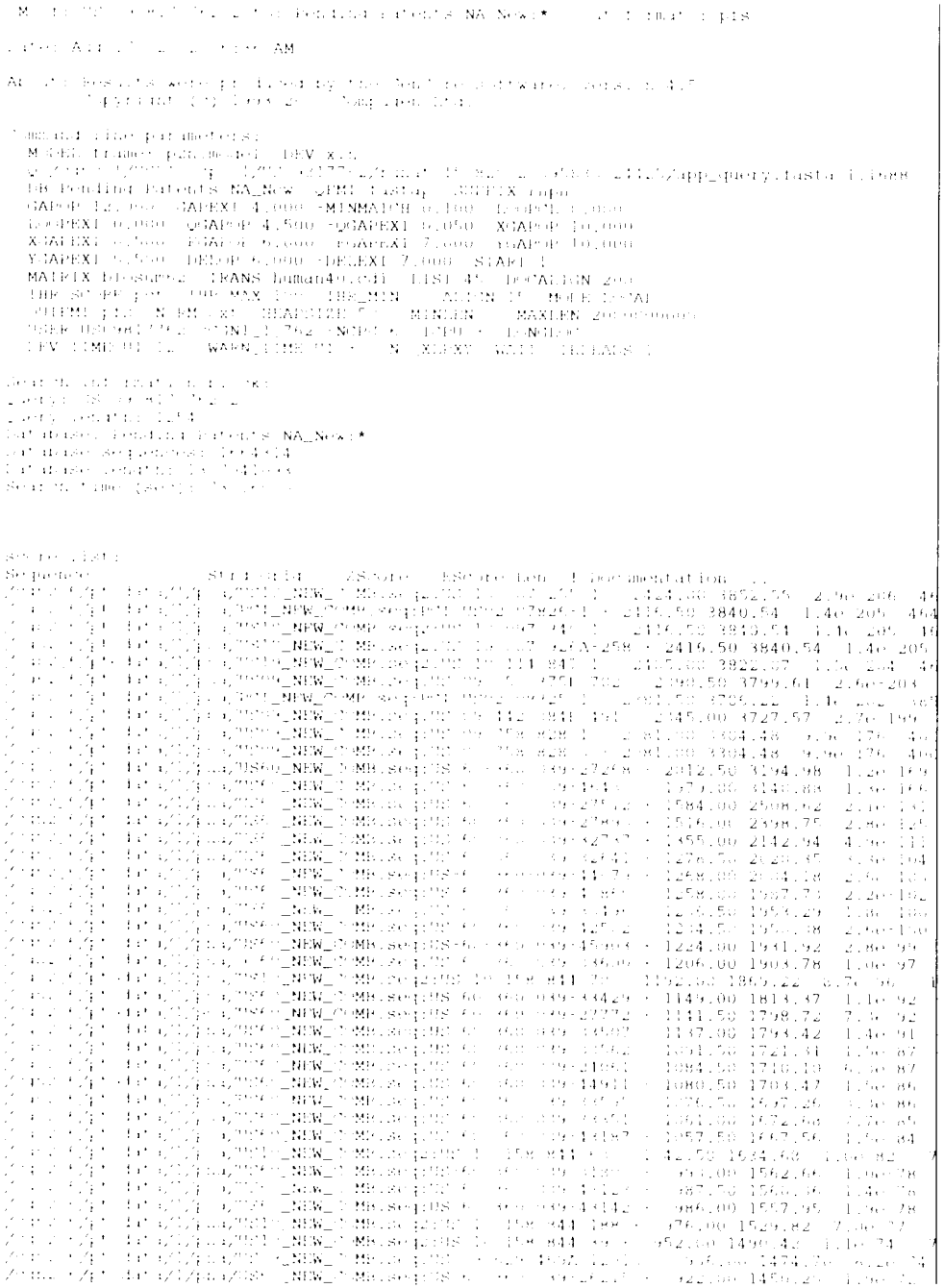




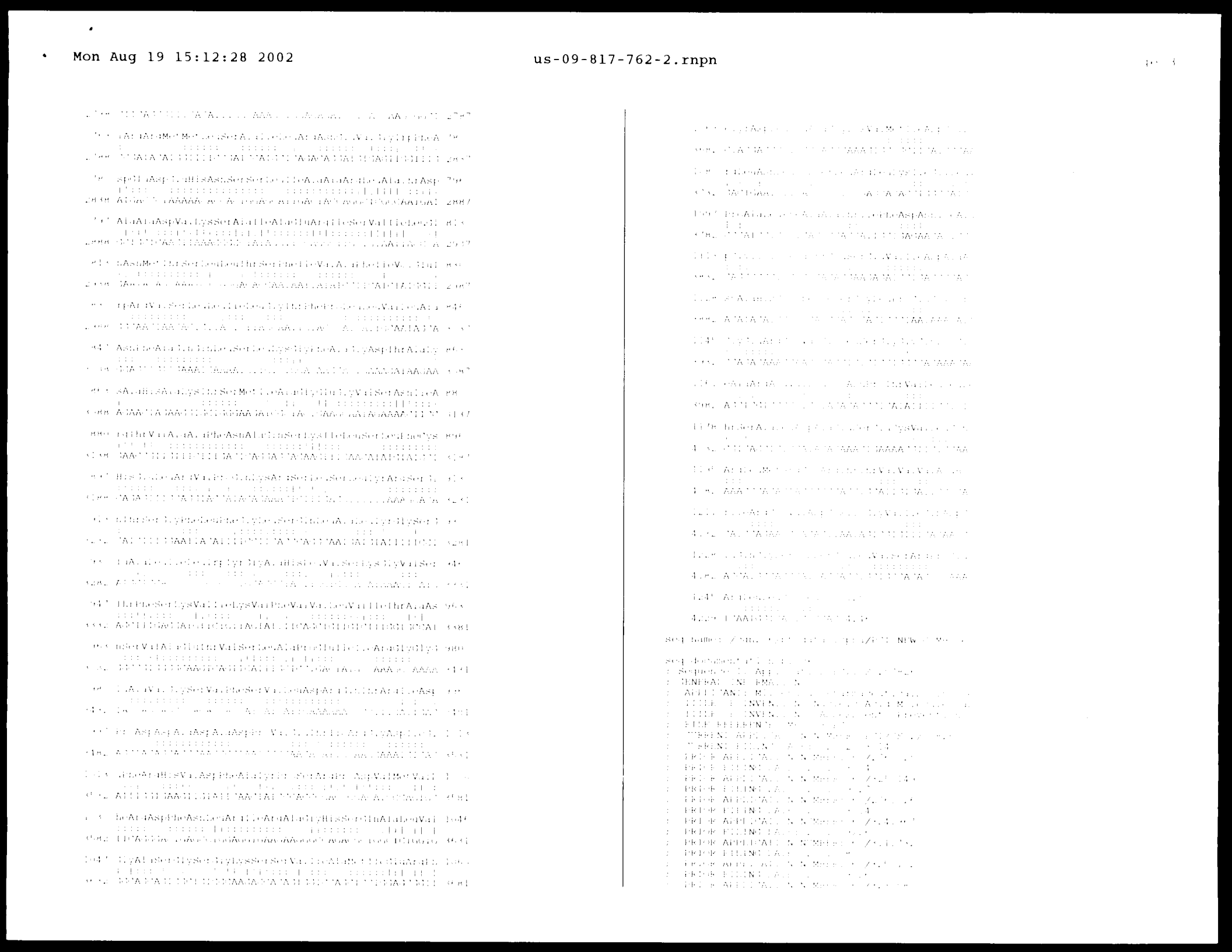


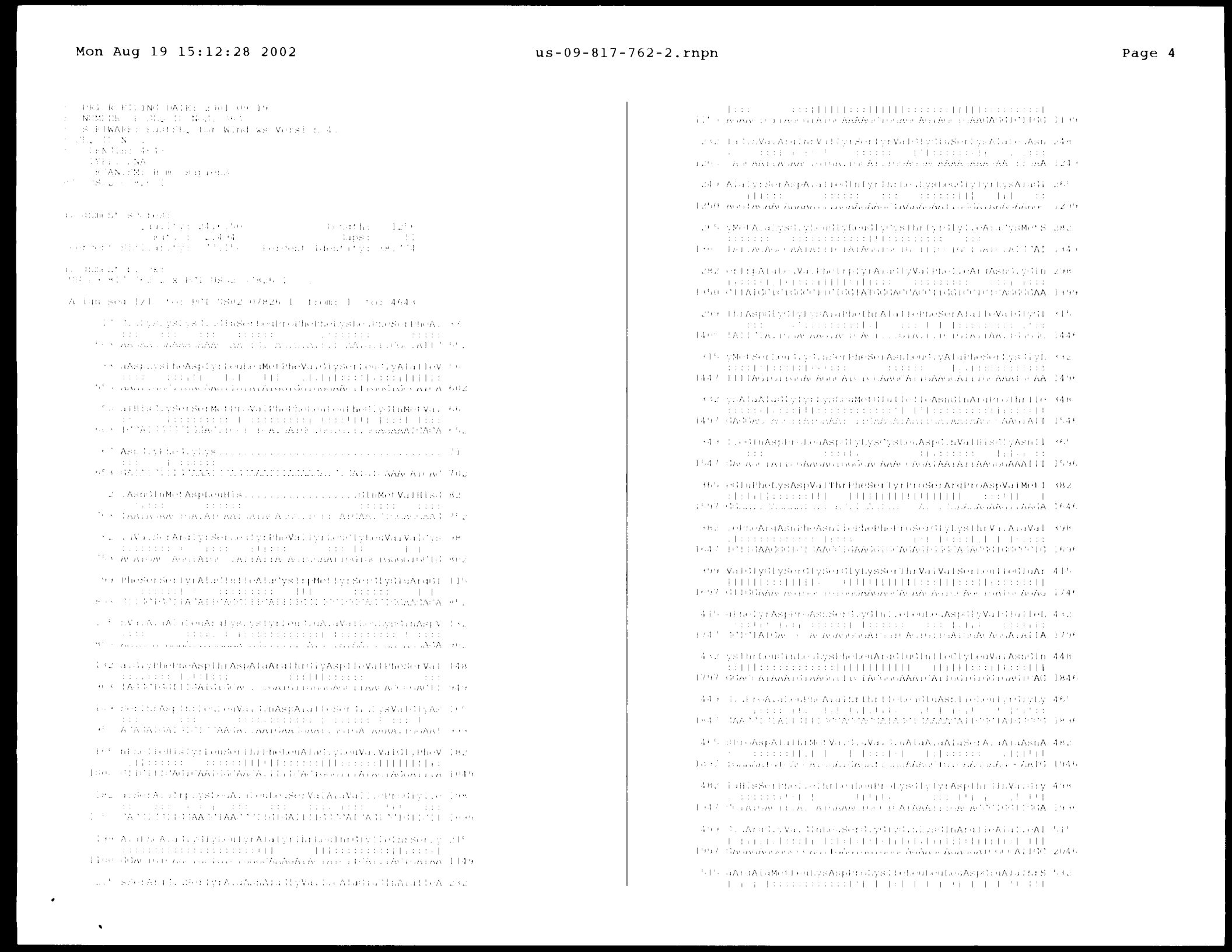


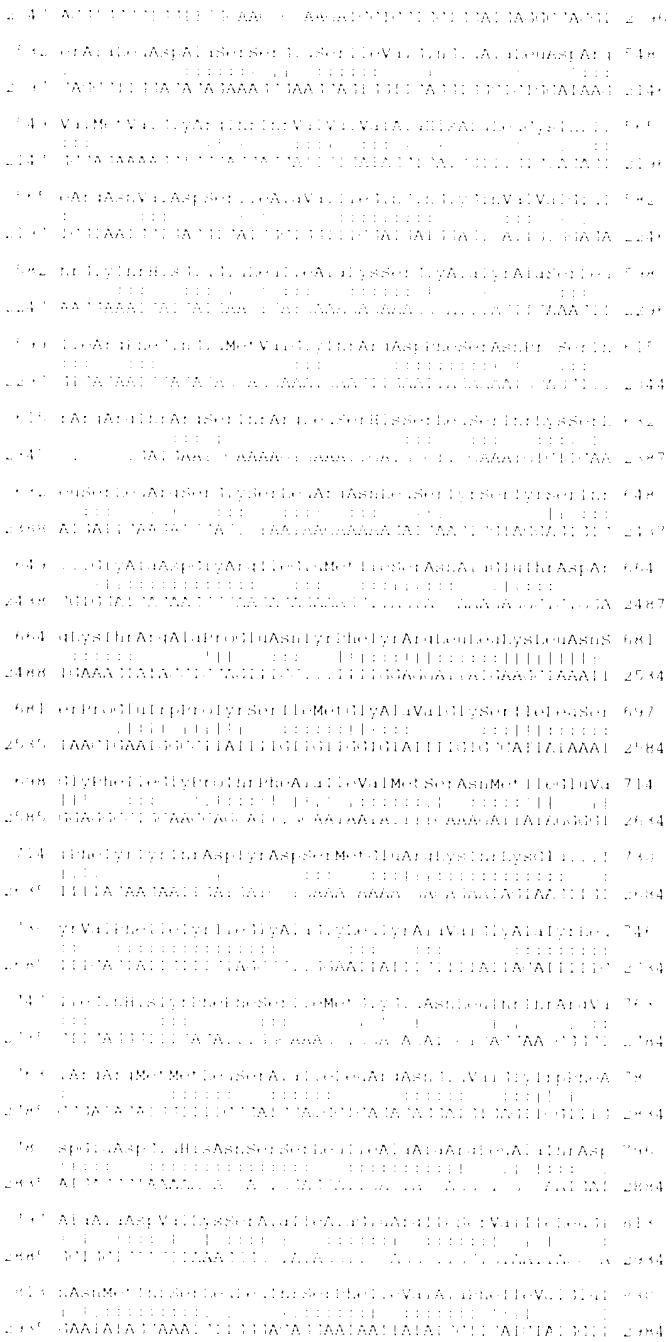


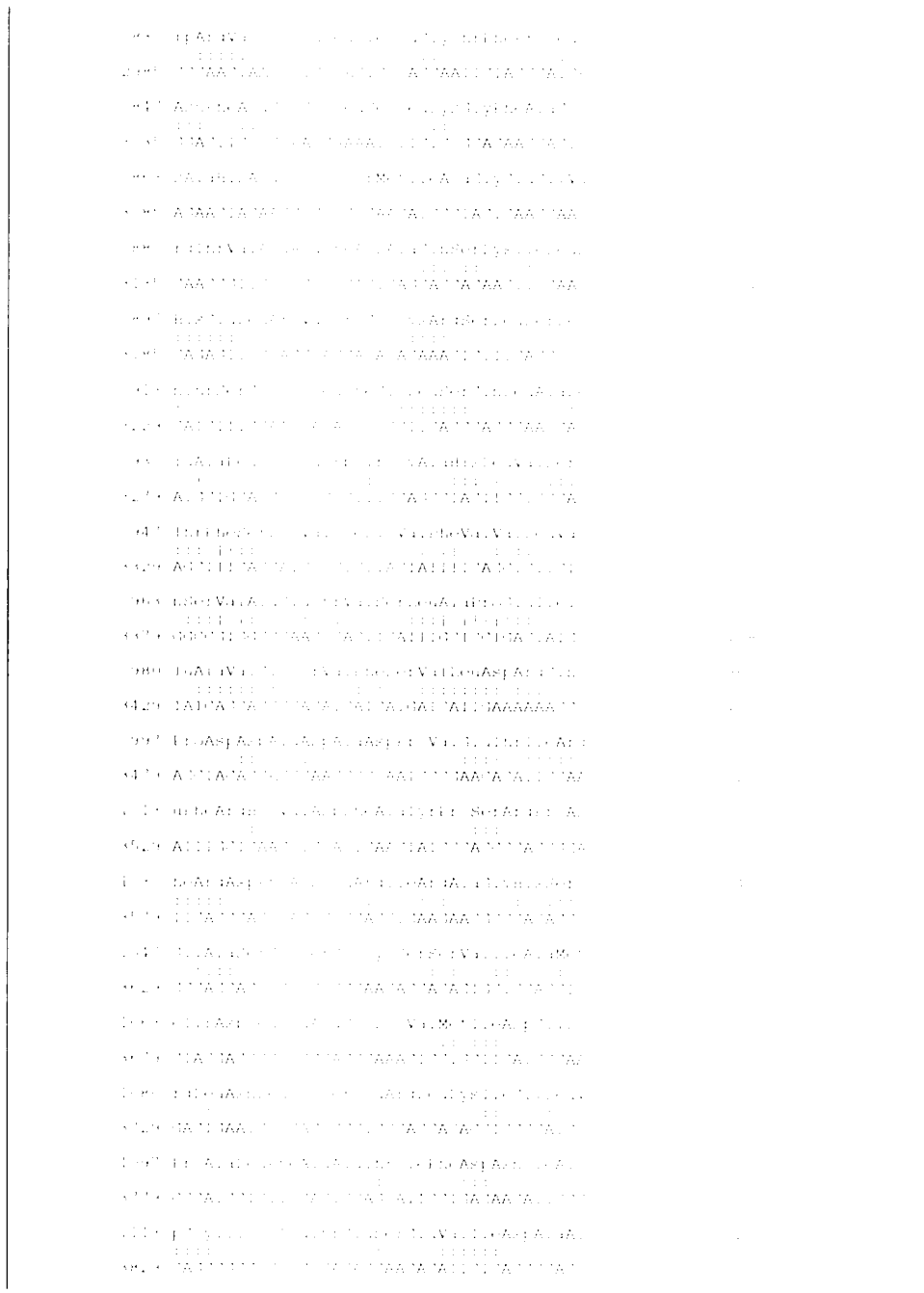


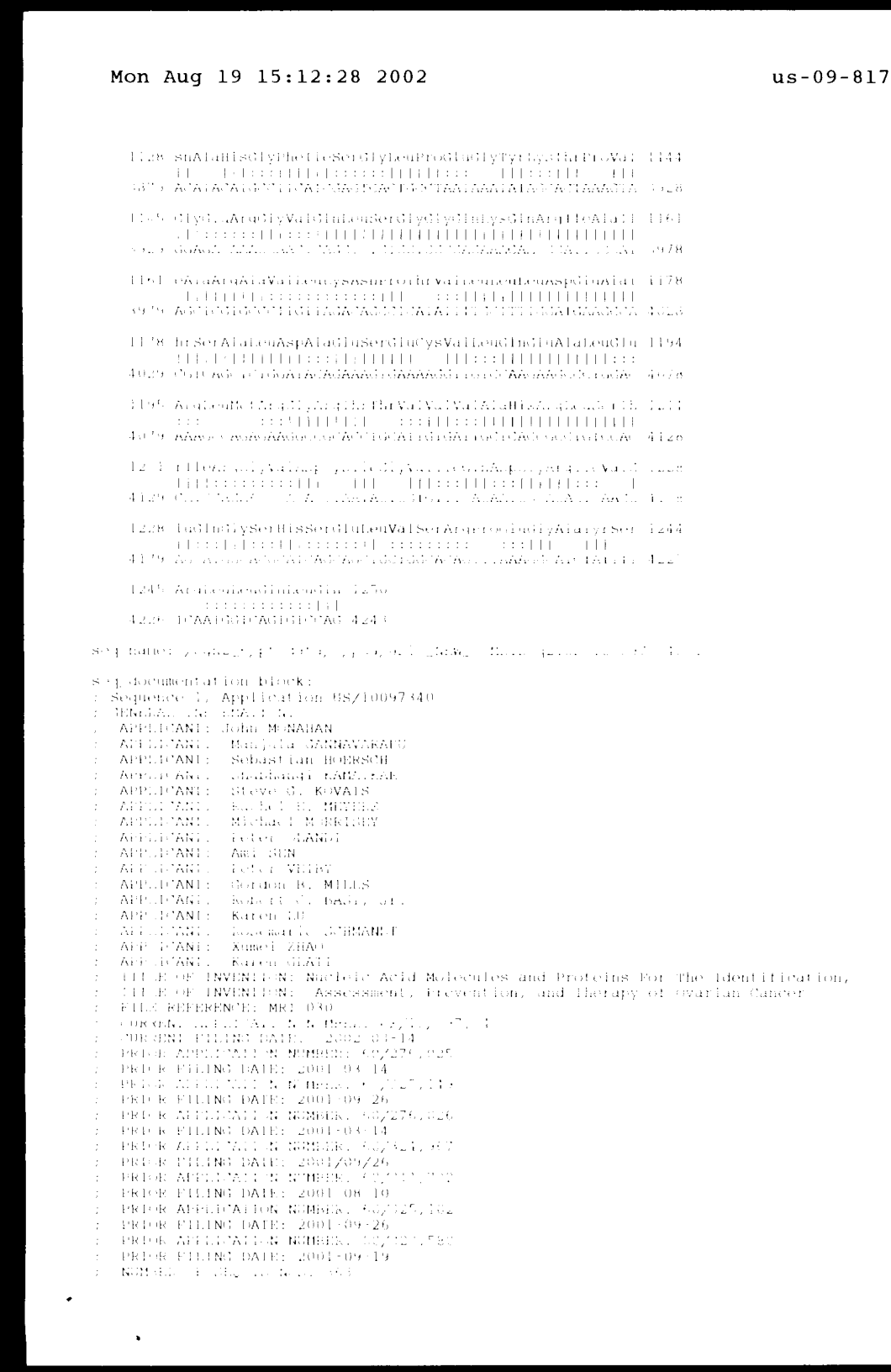












; SEQUENCE: PatentIn for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4644
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-097-340-1

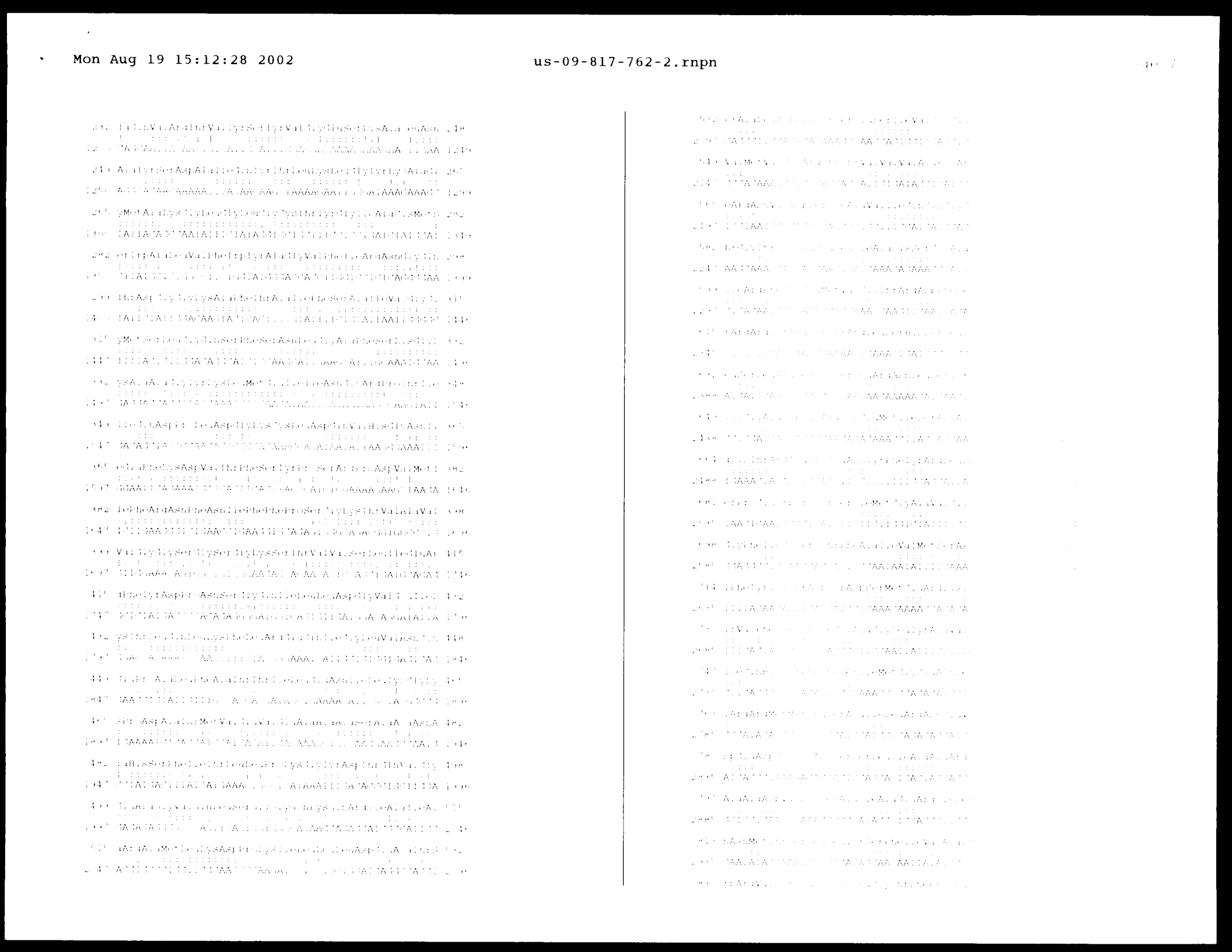
alignment scores:
Quality: 2416.50 Length: 1256
Ratio: 2.494 Gaps: 11
Percent Similarity: 77.150 Percent Identity: 68.774

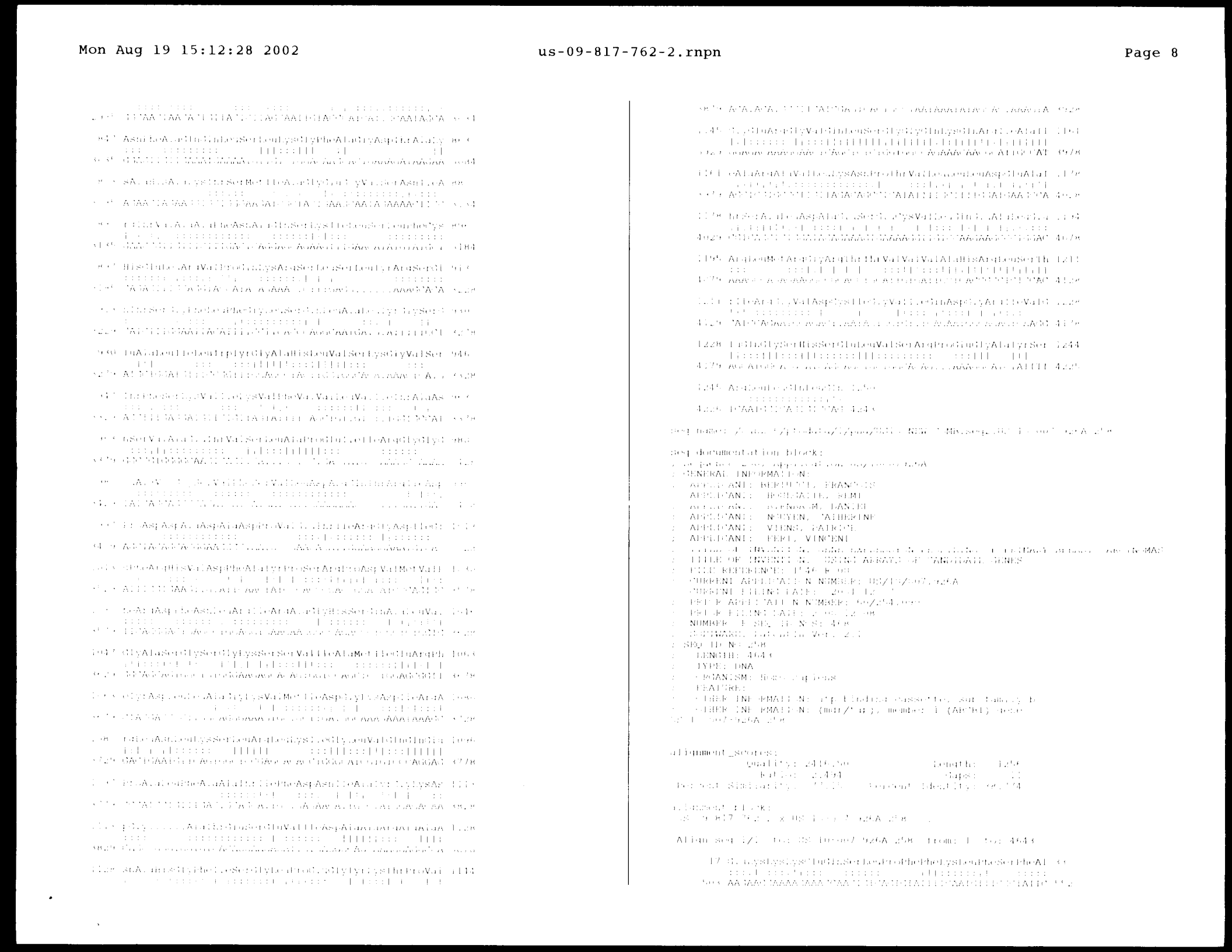
alignment block:
US-09-817-762-2 (1-1256) x US-10-097-340-1 ..

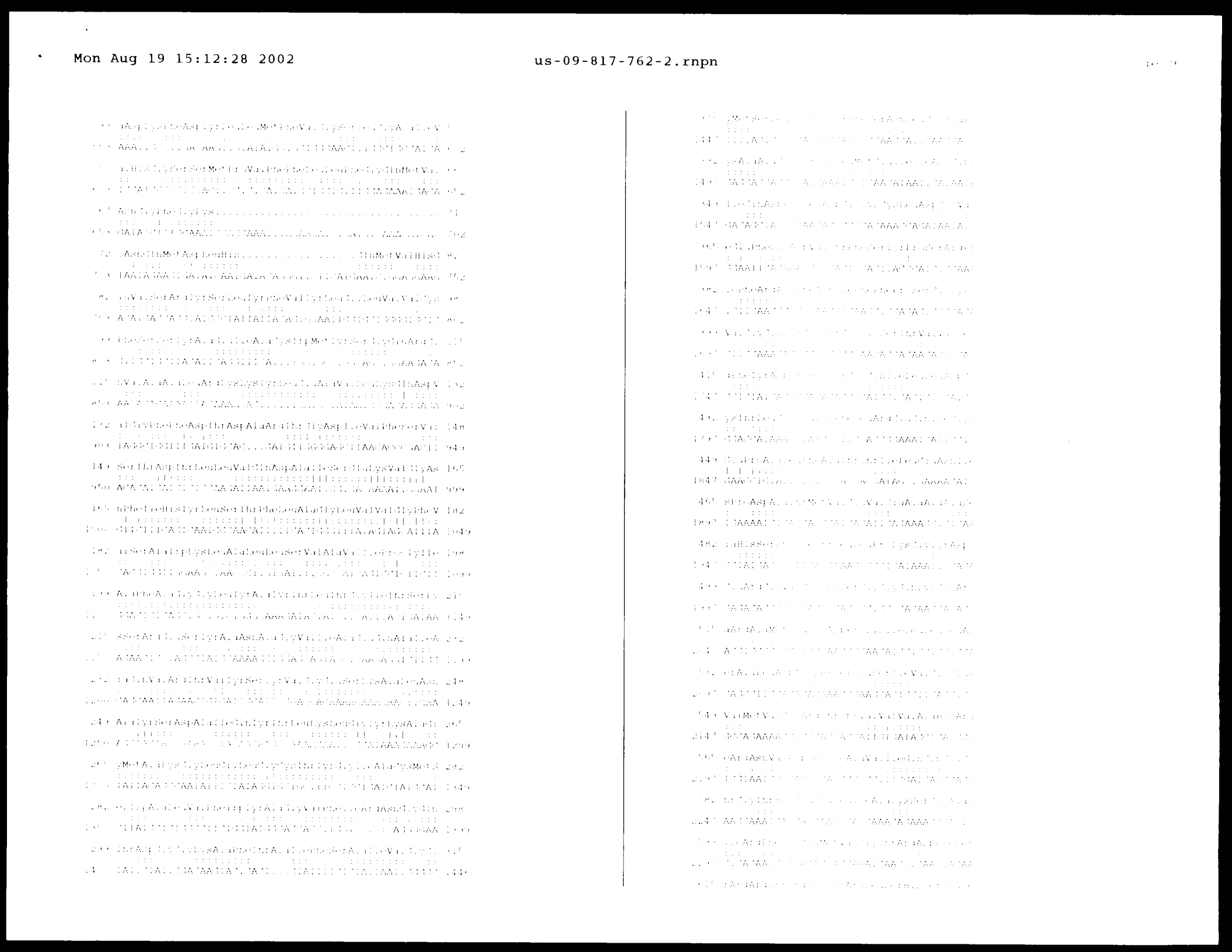
Alignment 1/1 from US-09-817-762-2 (1-1256) to US-10-097-340-1 (from 1 to 4644)

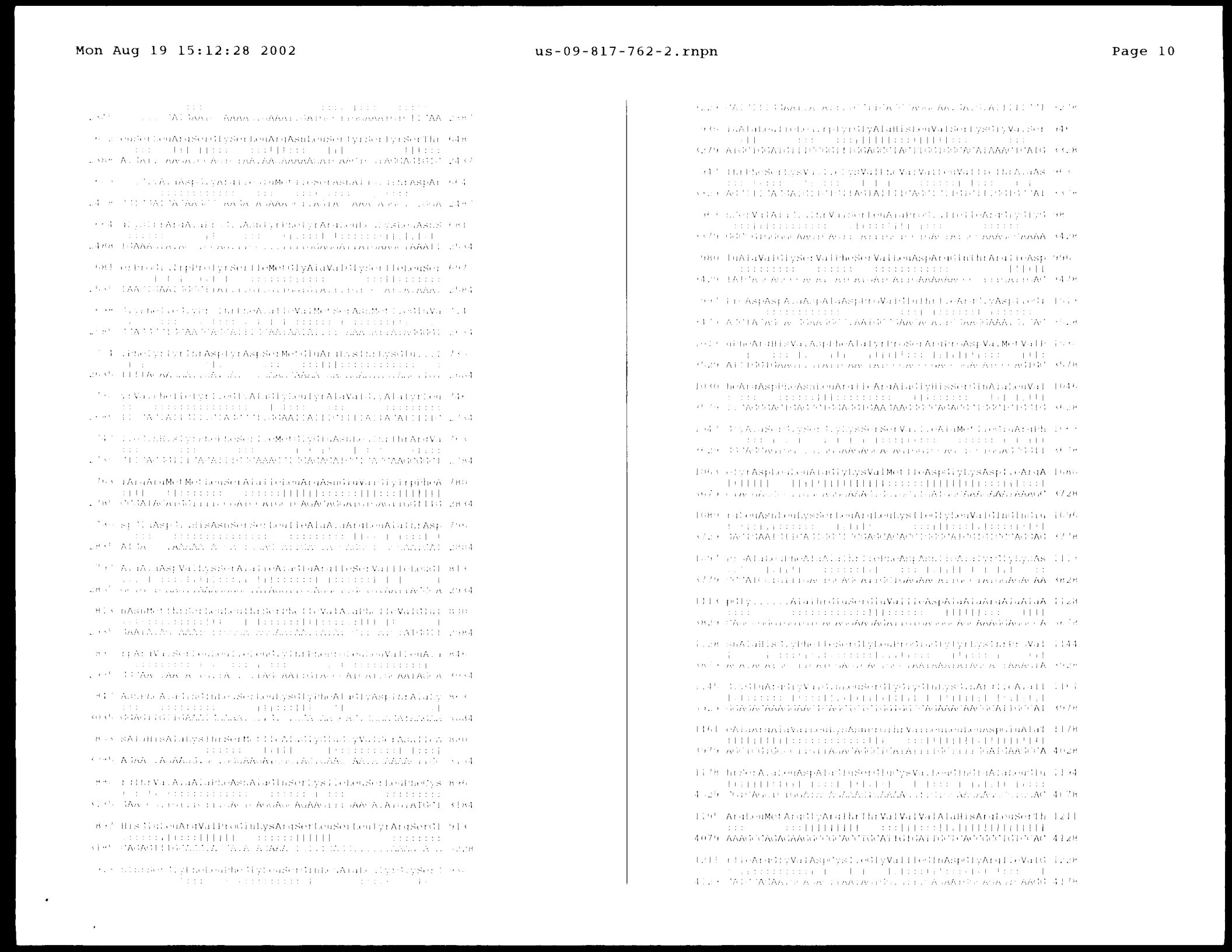
; SEQUENCE: PatentIn for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4644
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-097-340-1

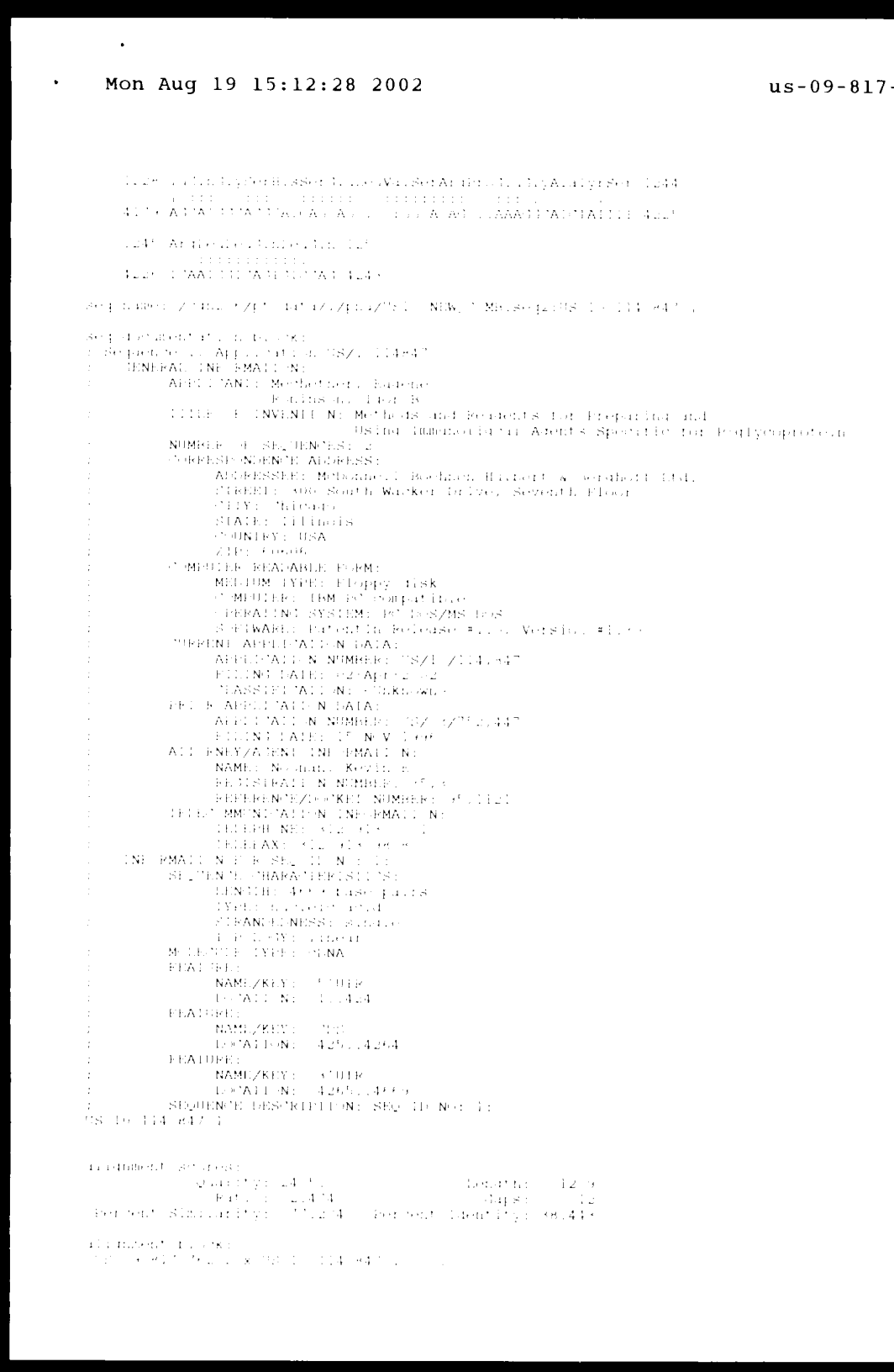
Sequence 1, Application US/10097340
; PATENT NO: US20020137046A1
; APPLICANT: John MENAHAN
; APPLICANT: Manjula GANNAVARAM
; APPLICANT: Sebastian BOERSCH
; APPLICANT: Shobhana KADAMBE
; APPLICANT: Steve G. KOVATS
; APPLICANT: Barbara B. BRITTEN
; APPLICANT: Michael M BRITTEN
; APPLICANT: Peter GANGI
; APPLICANT: Amy REN
; APPLICANT: Peter VETTER
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, JR.
; APPLICANT: Karen LU
; APPLICANT: Anders ZIMMER
; APPLICANT: Xuewei ZHAO
; APPLICANT: Karen GLAFF
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI 030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,214
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/321,957
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/299,757
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: 60/325,192
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/325,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ IDS: 93

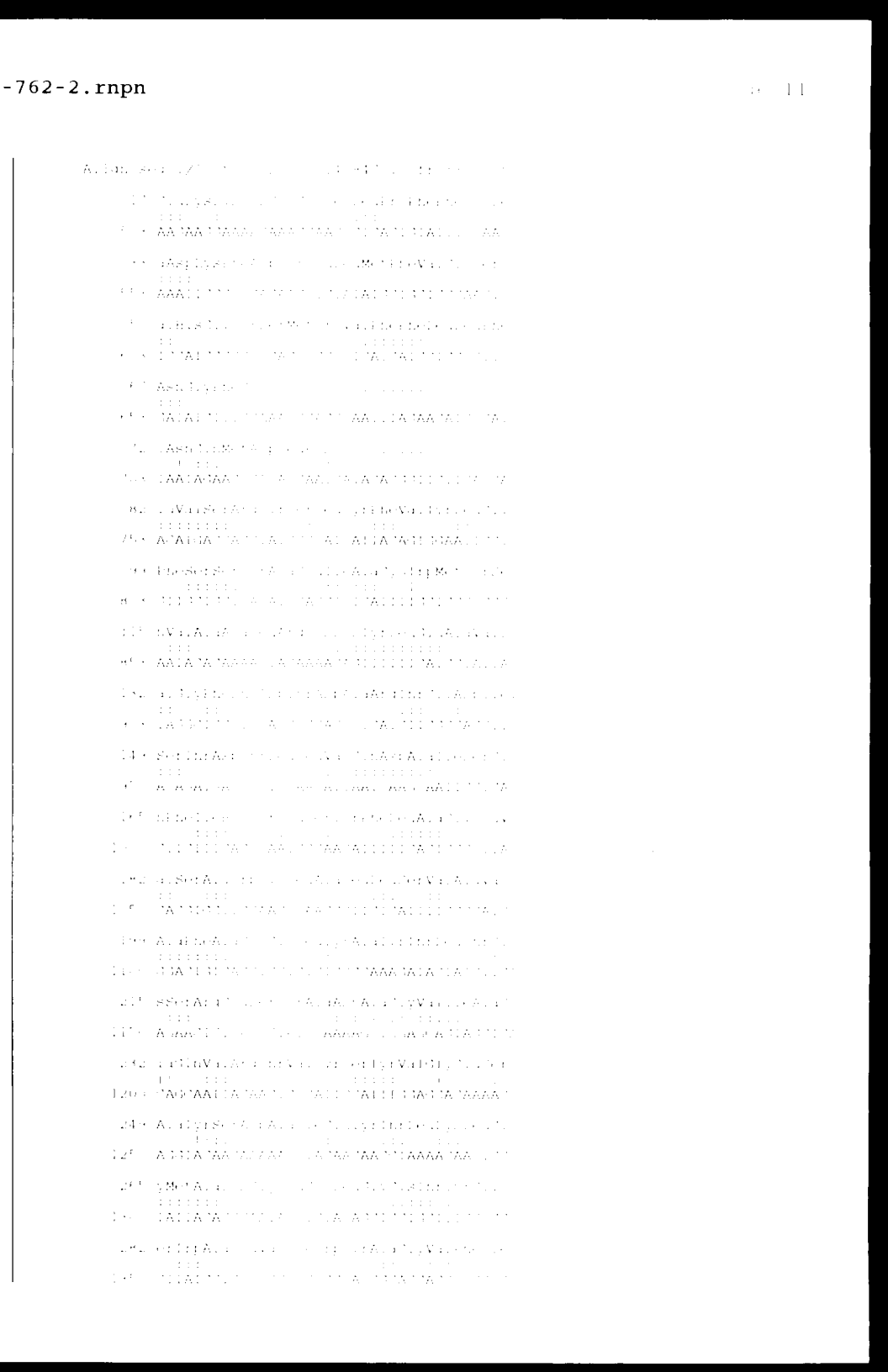


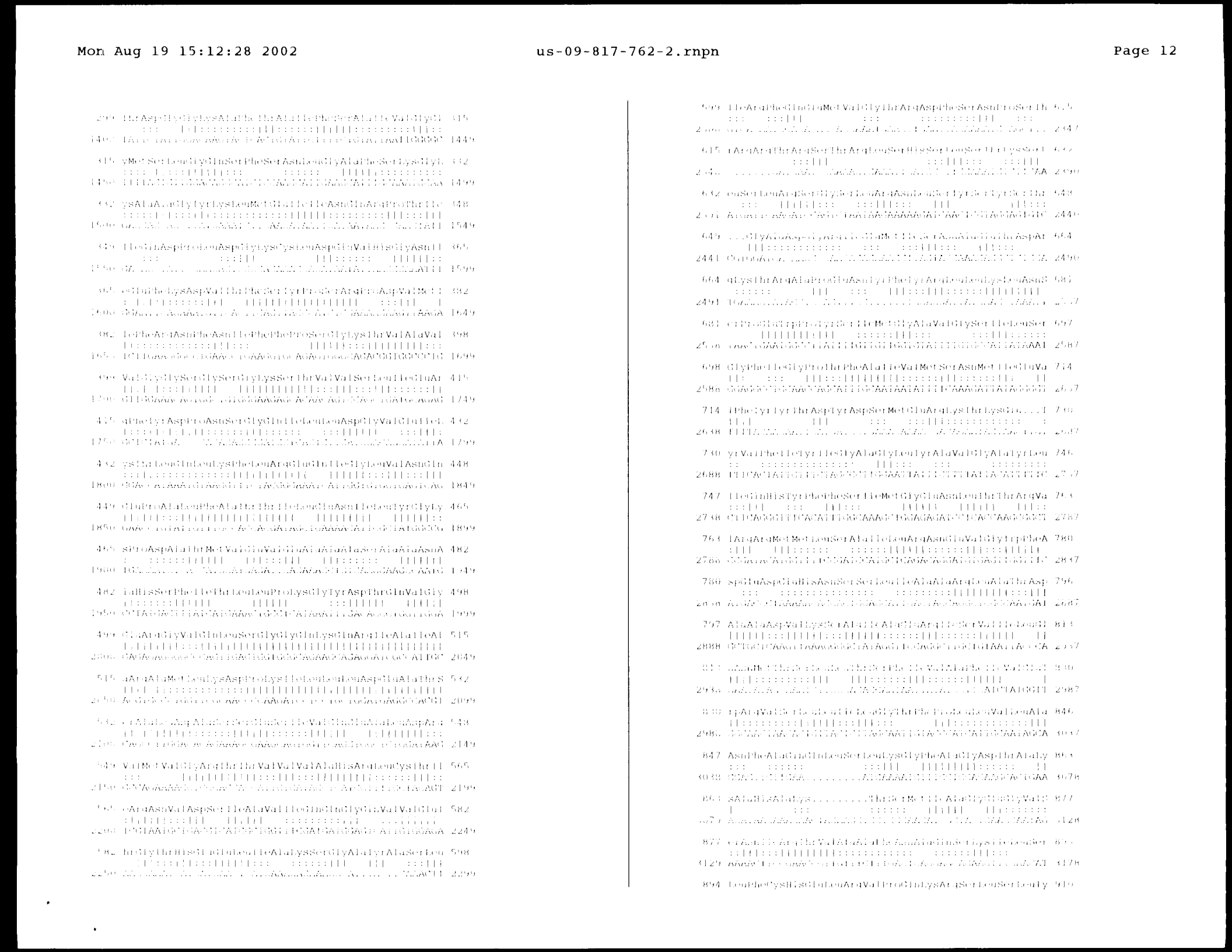


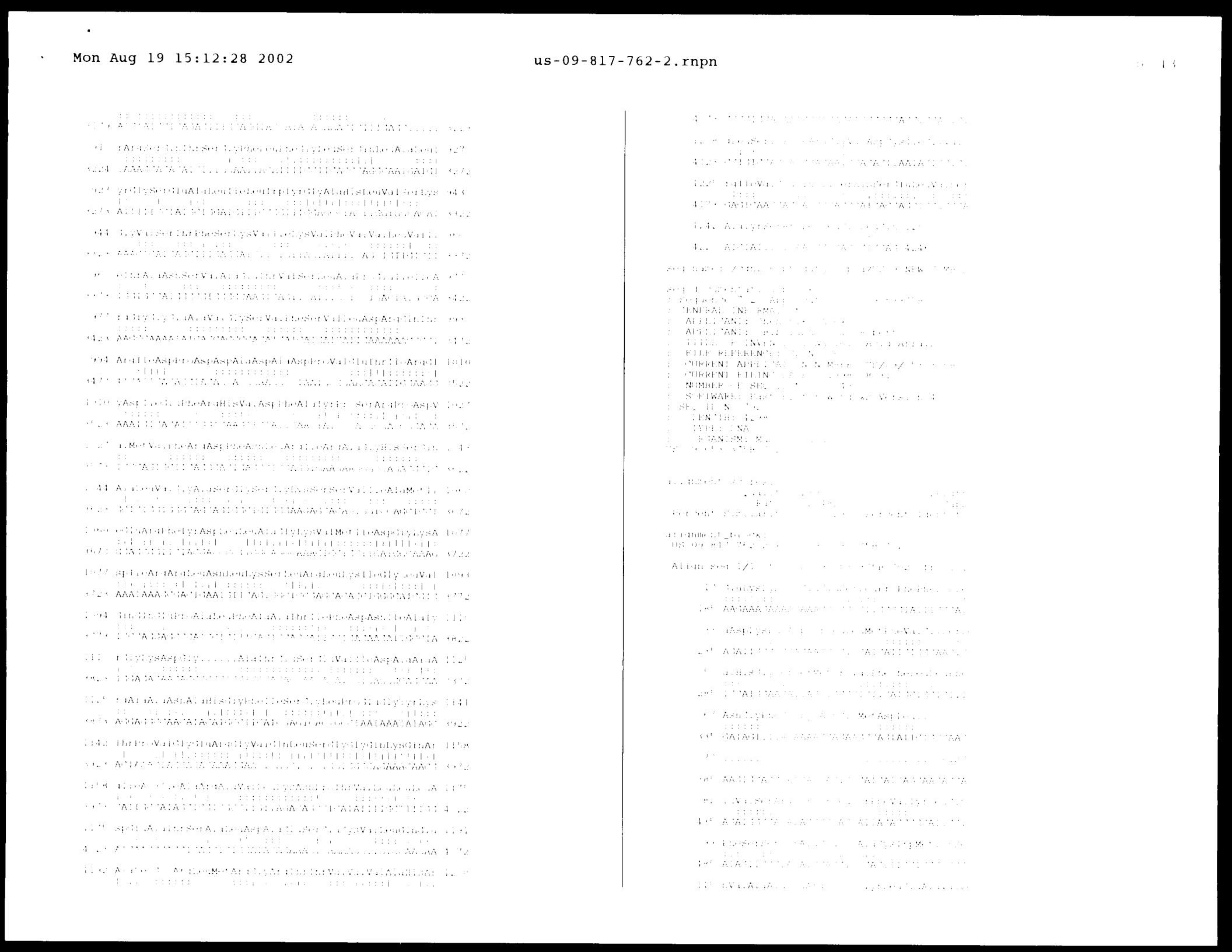


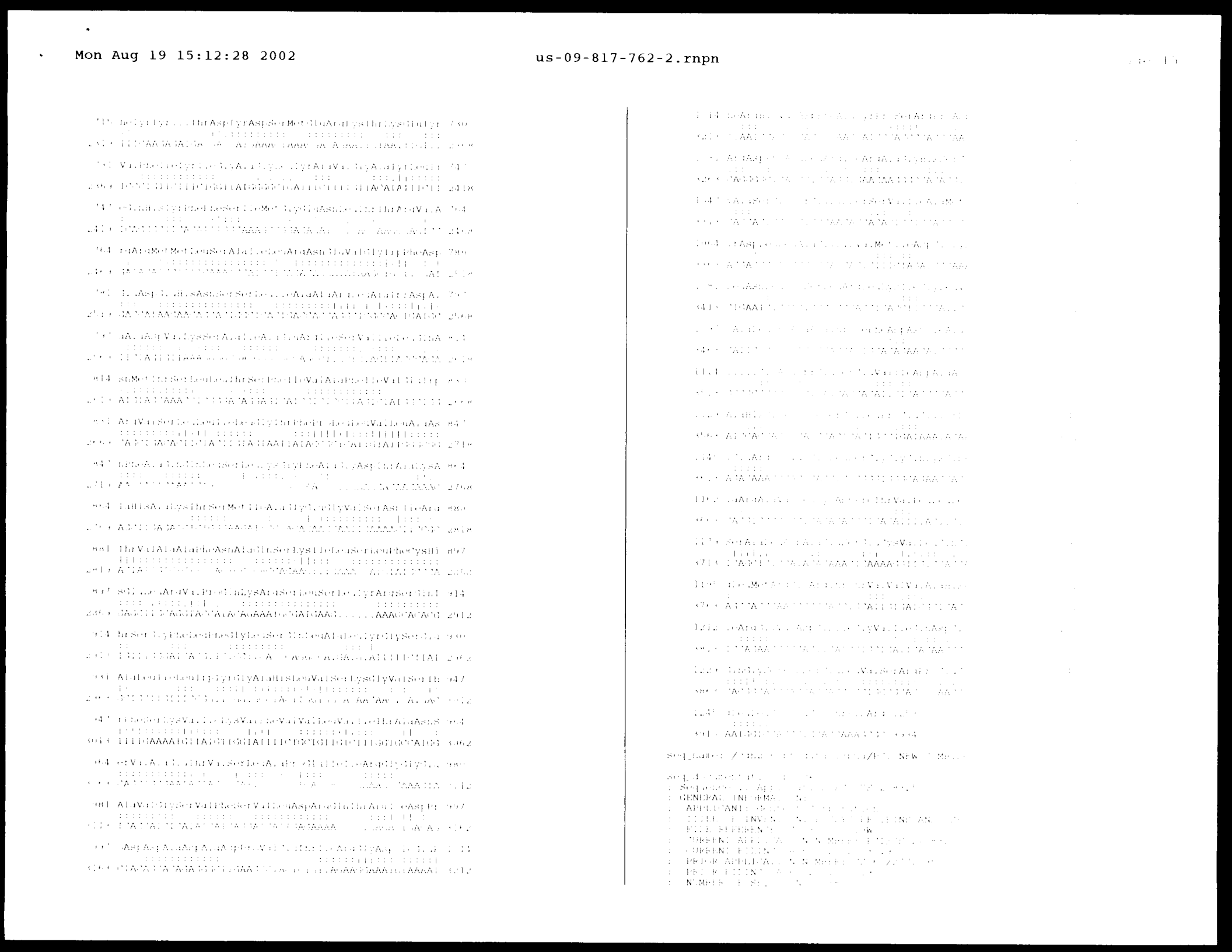


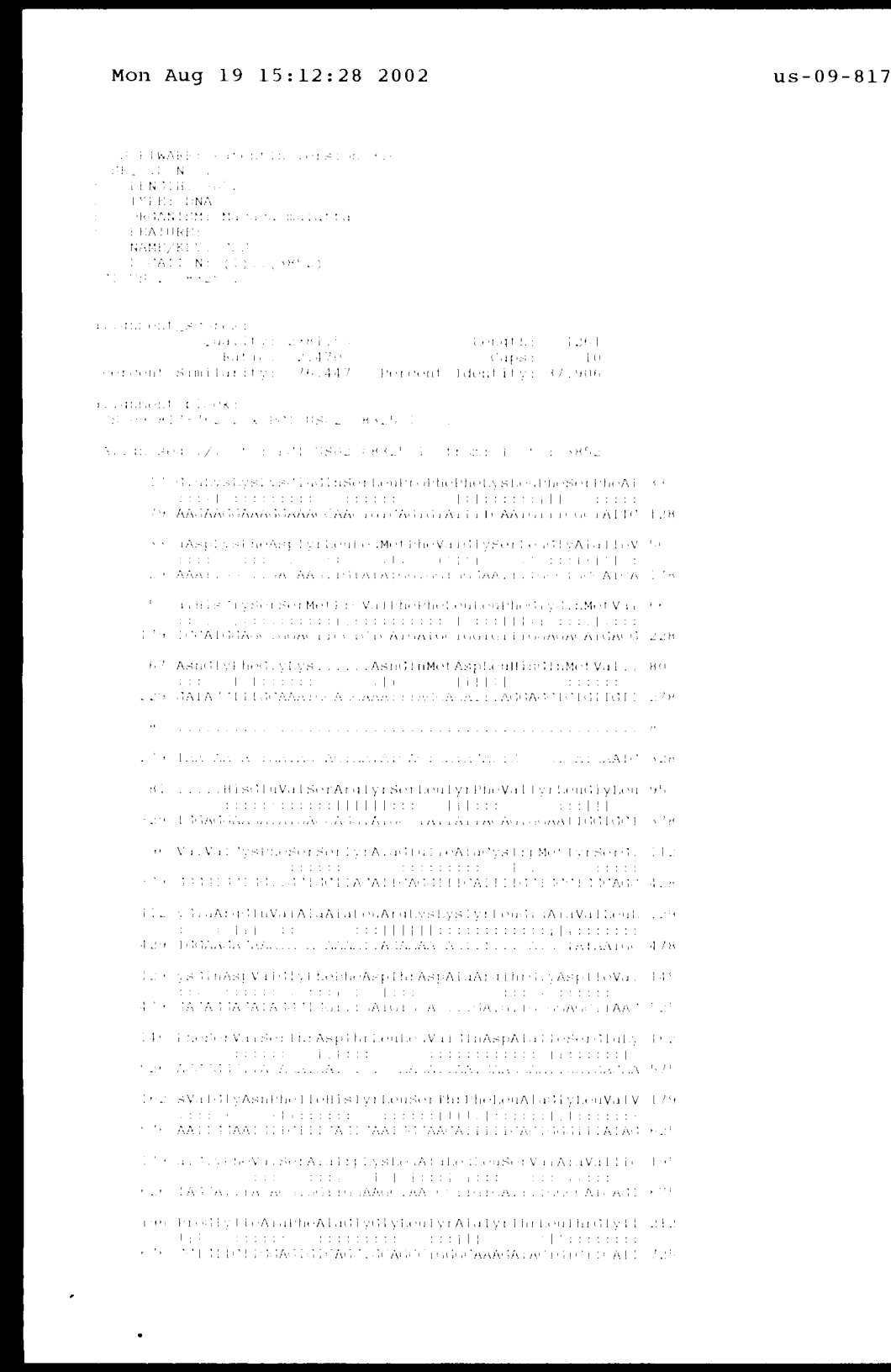


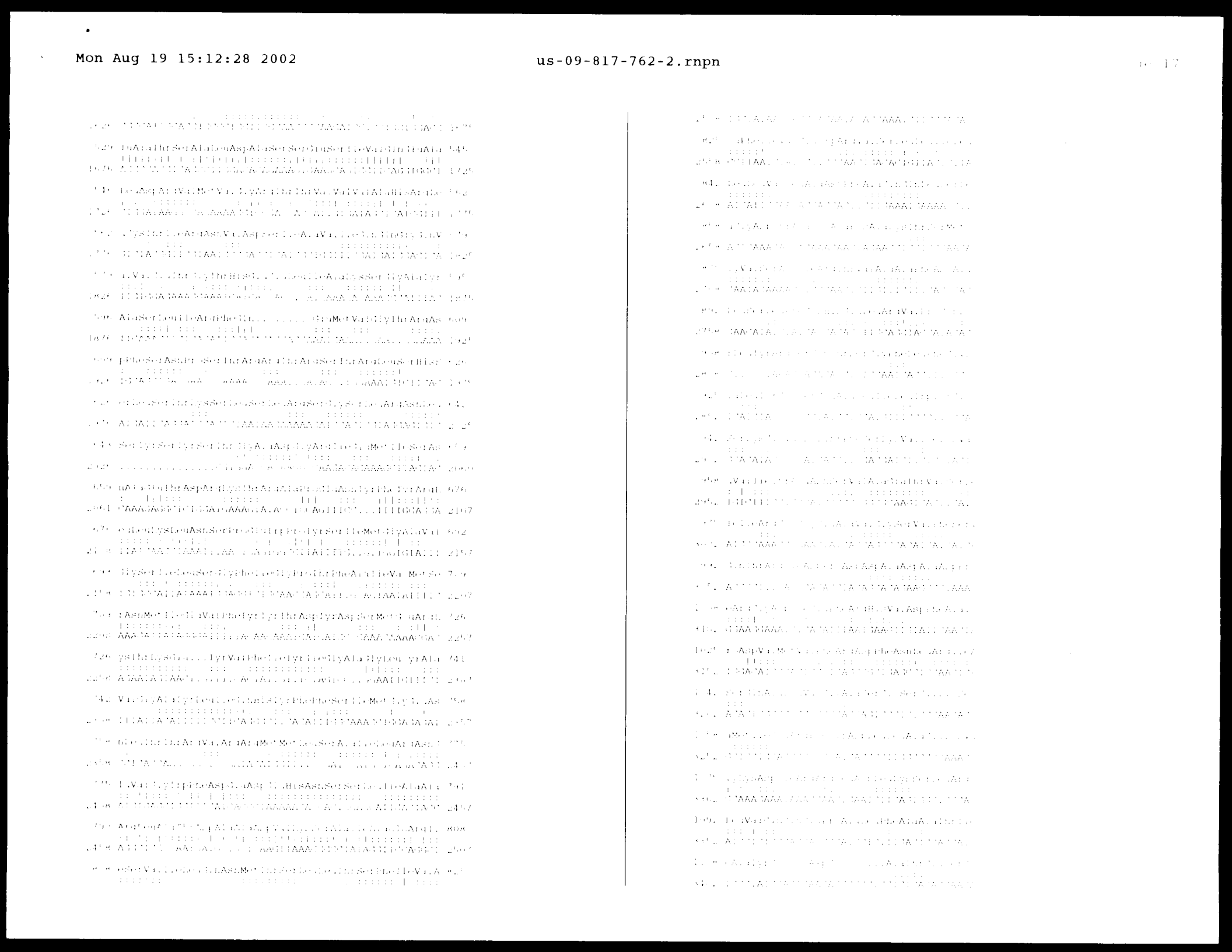


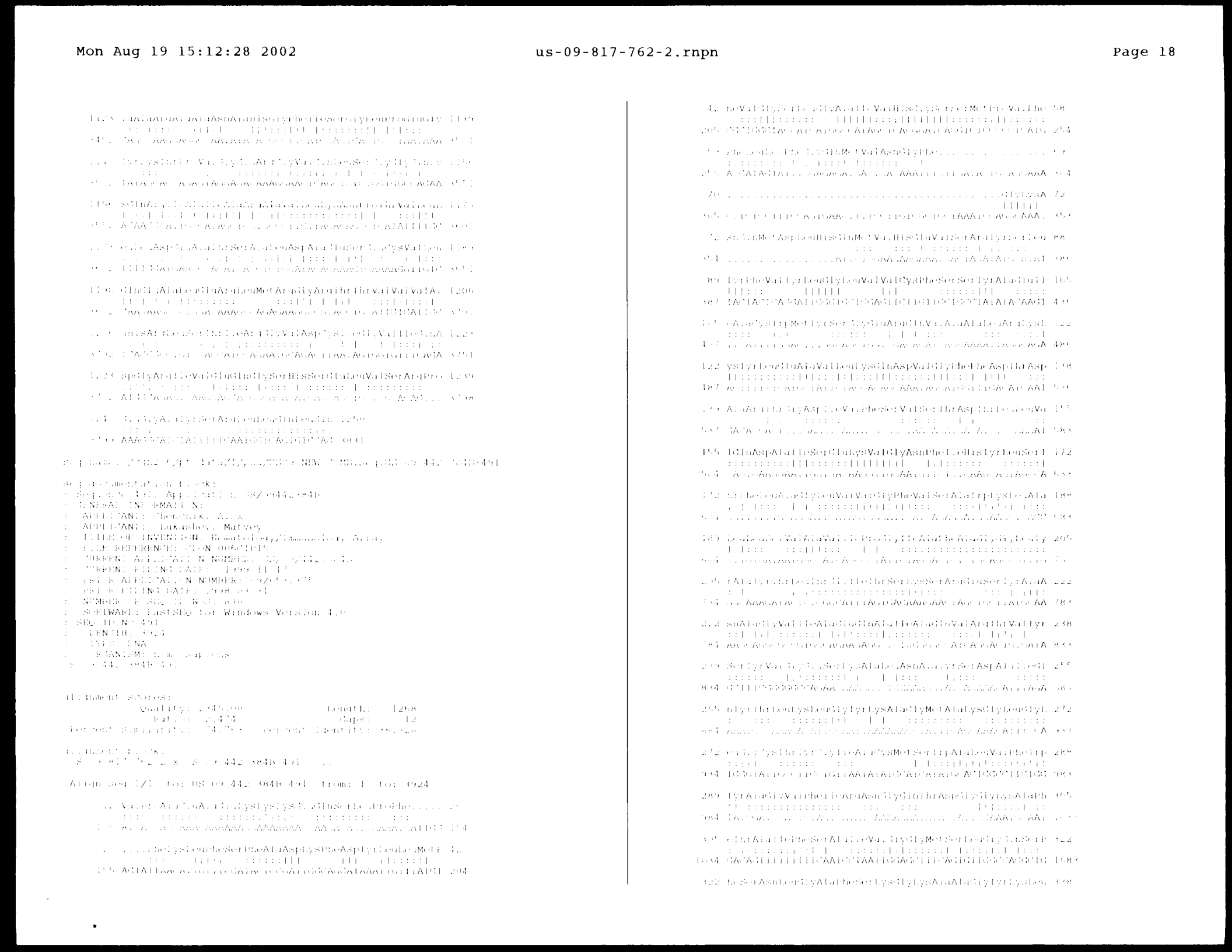


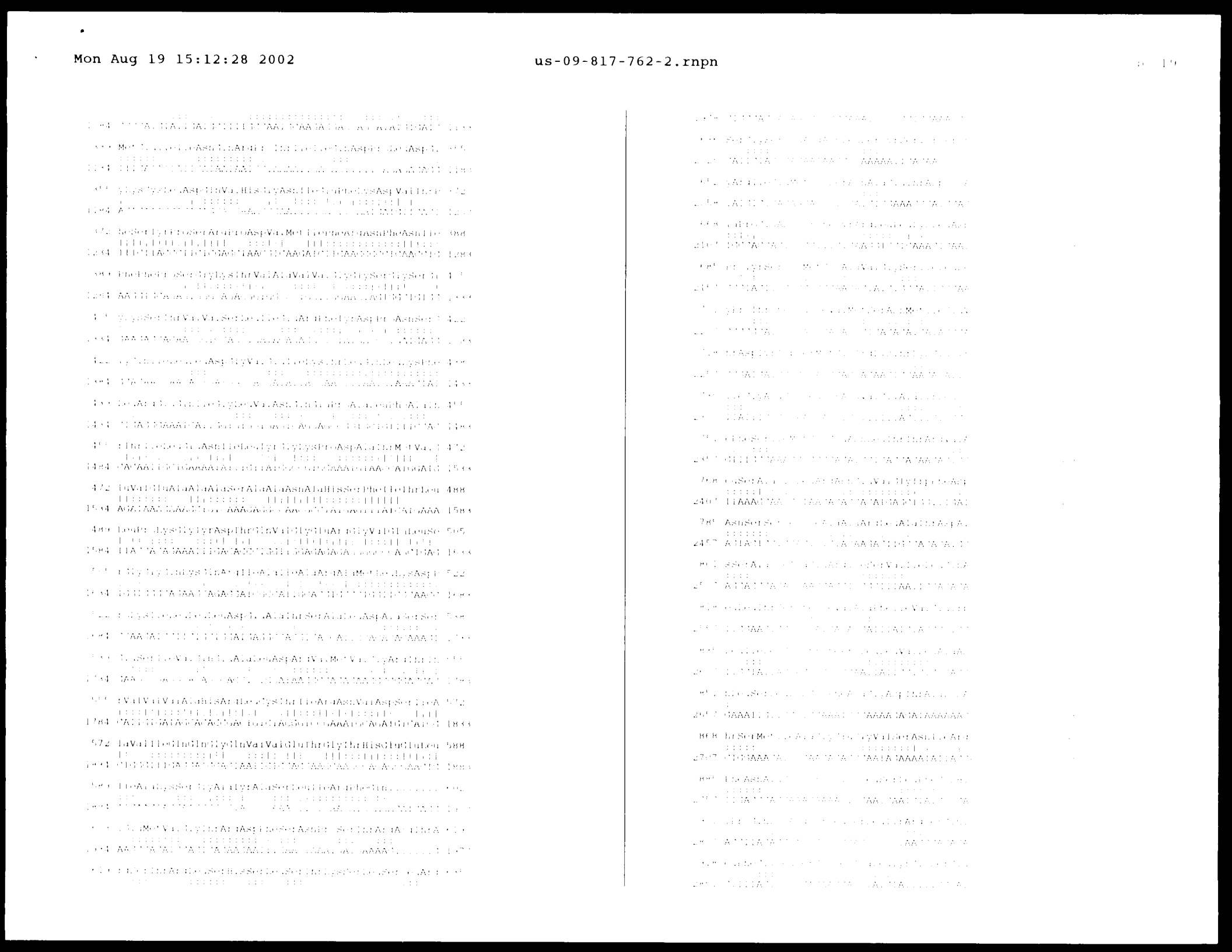


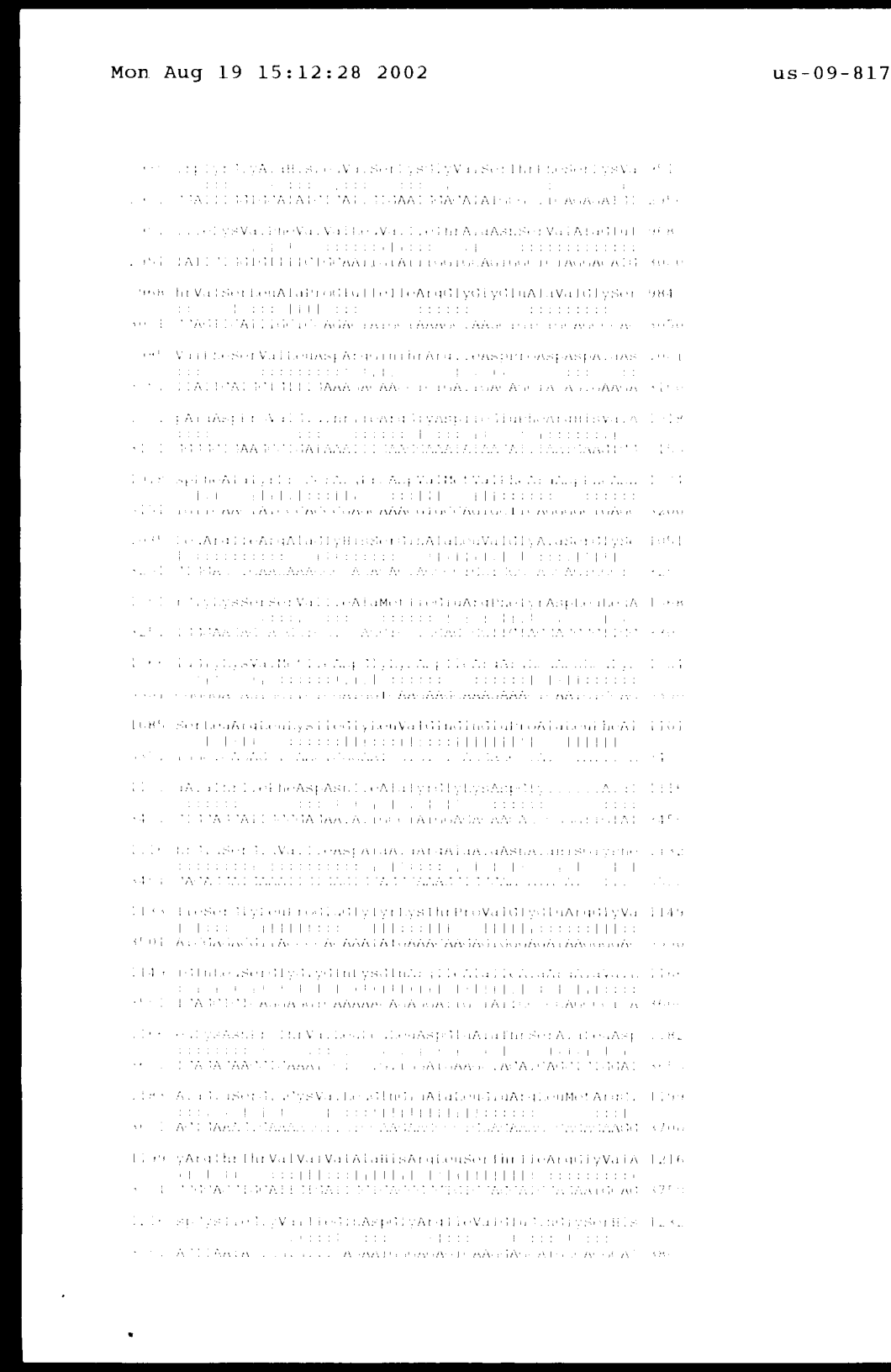


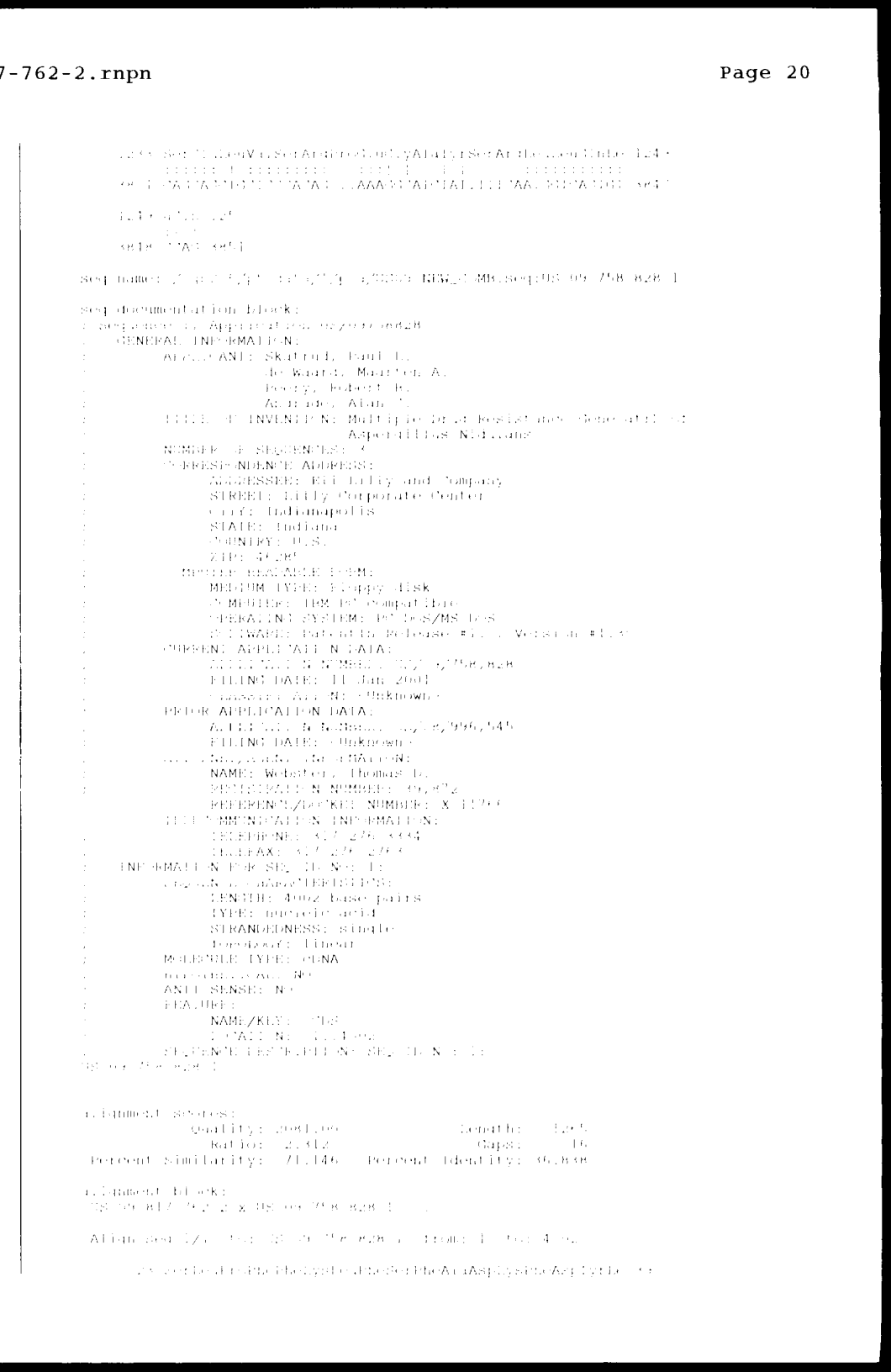


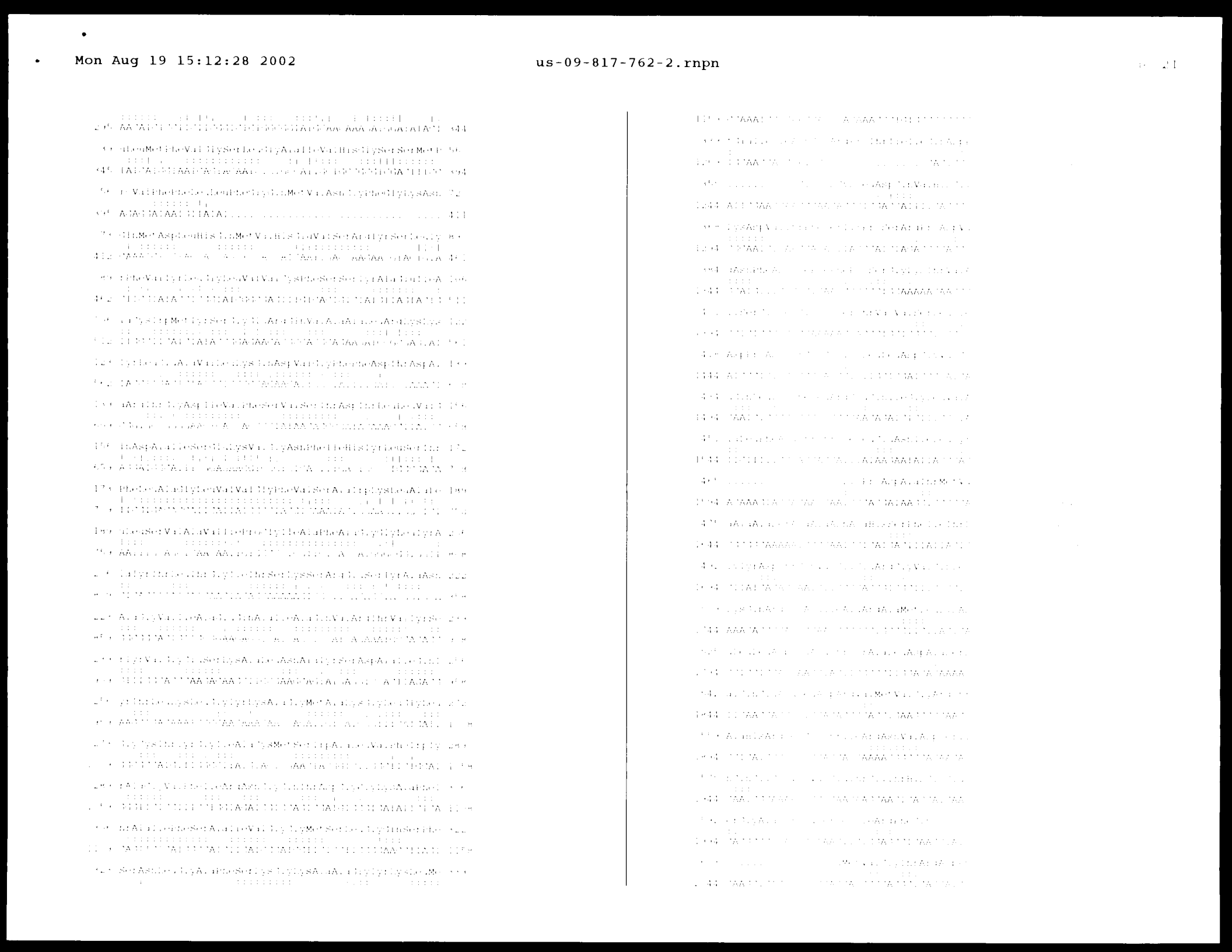


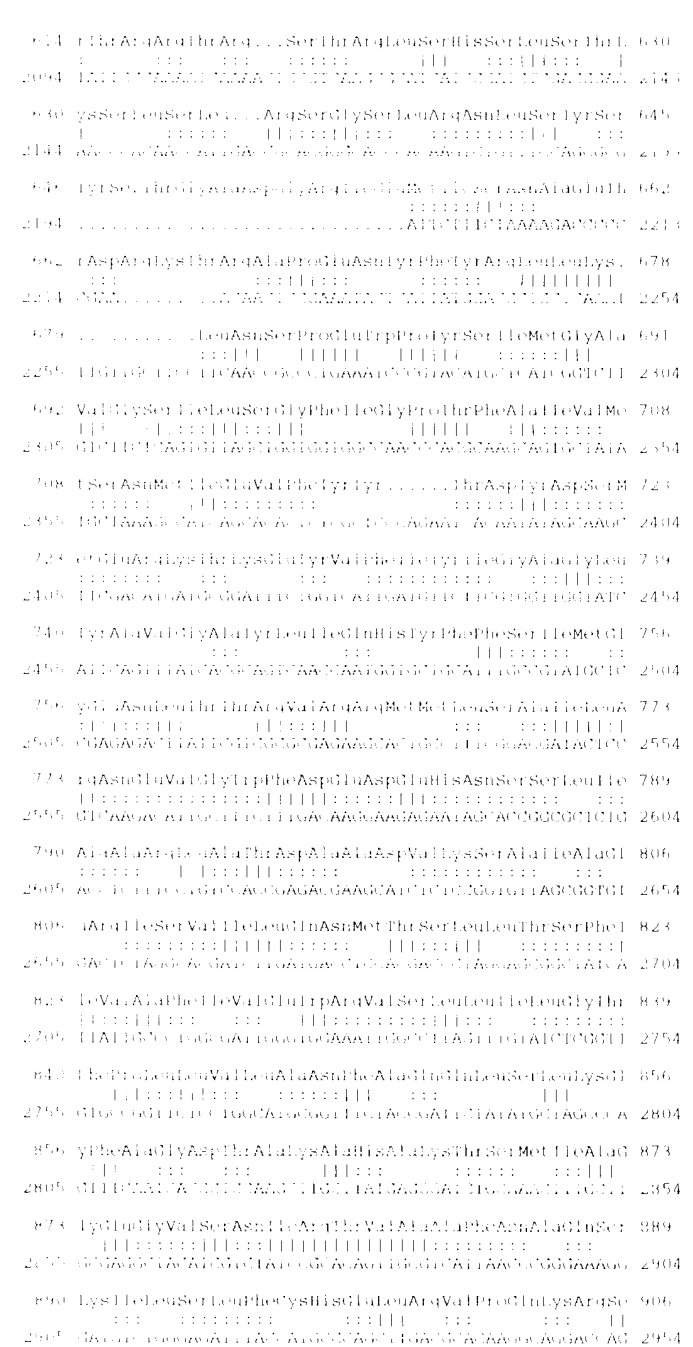


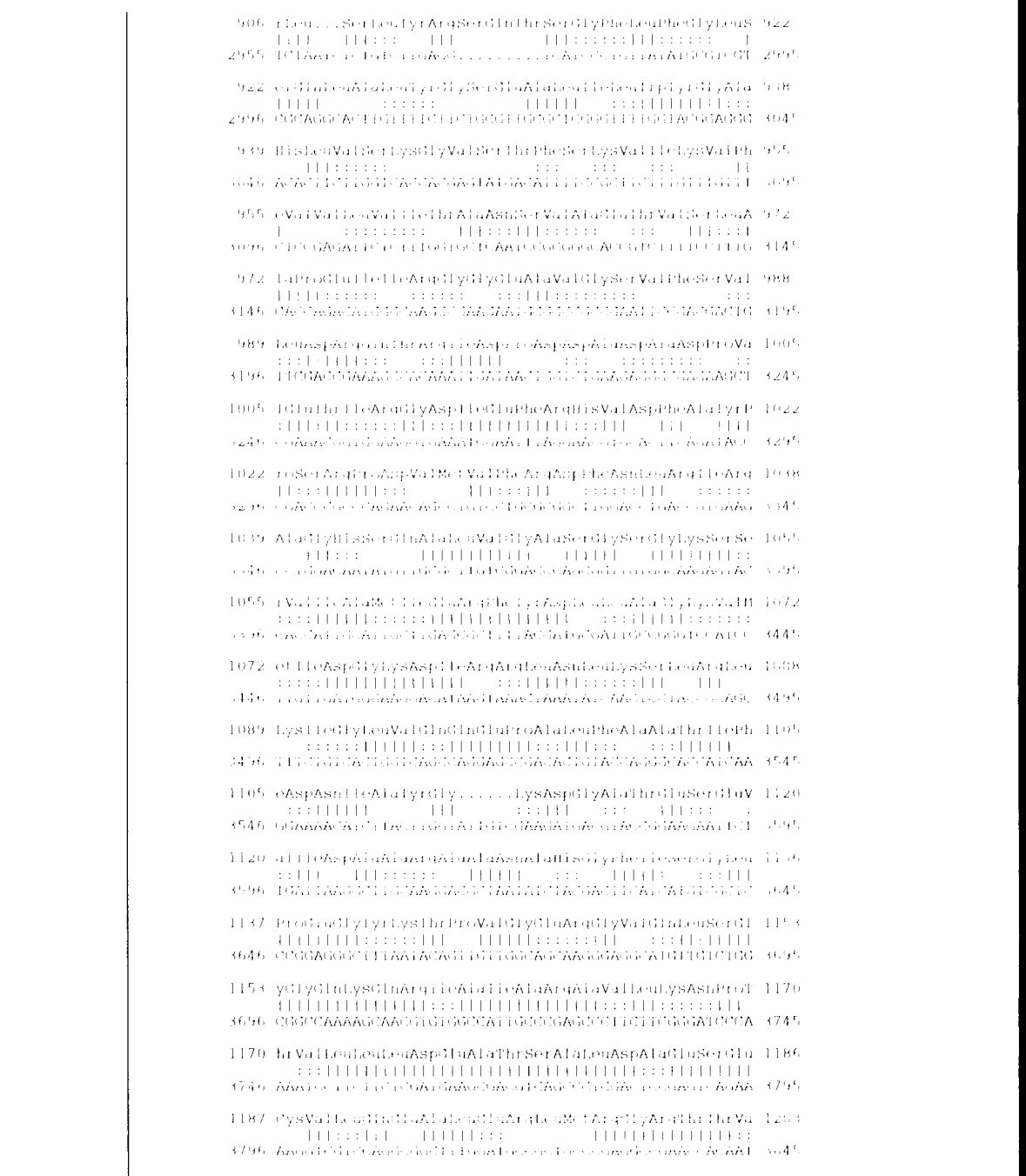


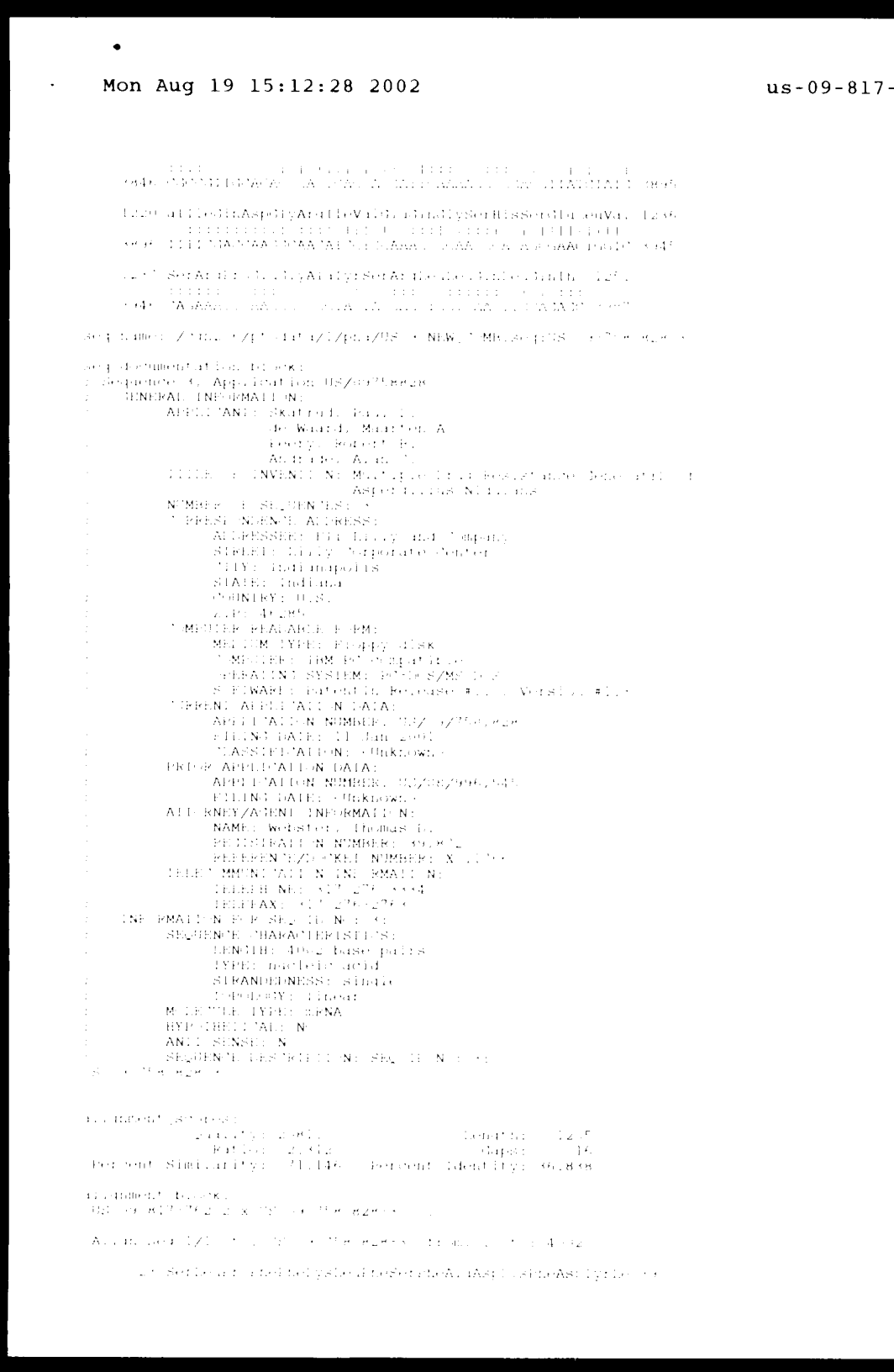


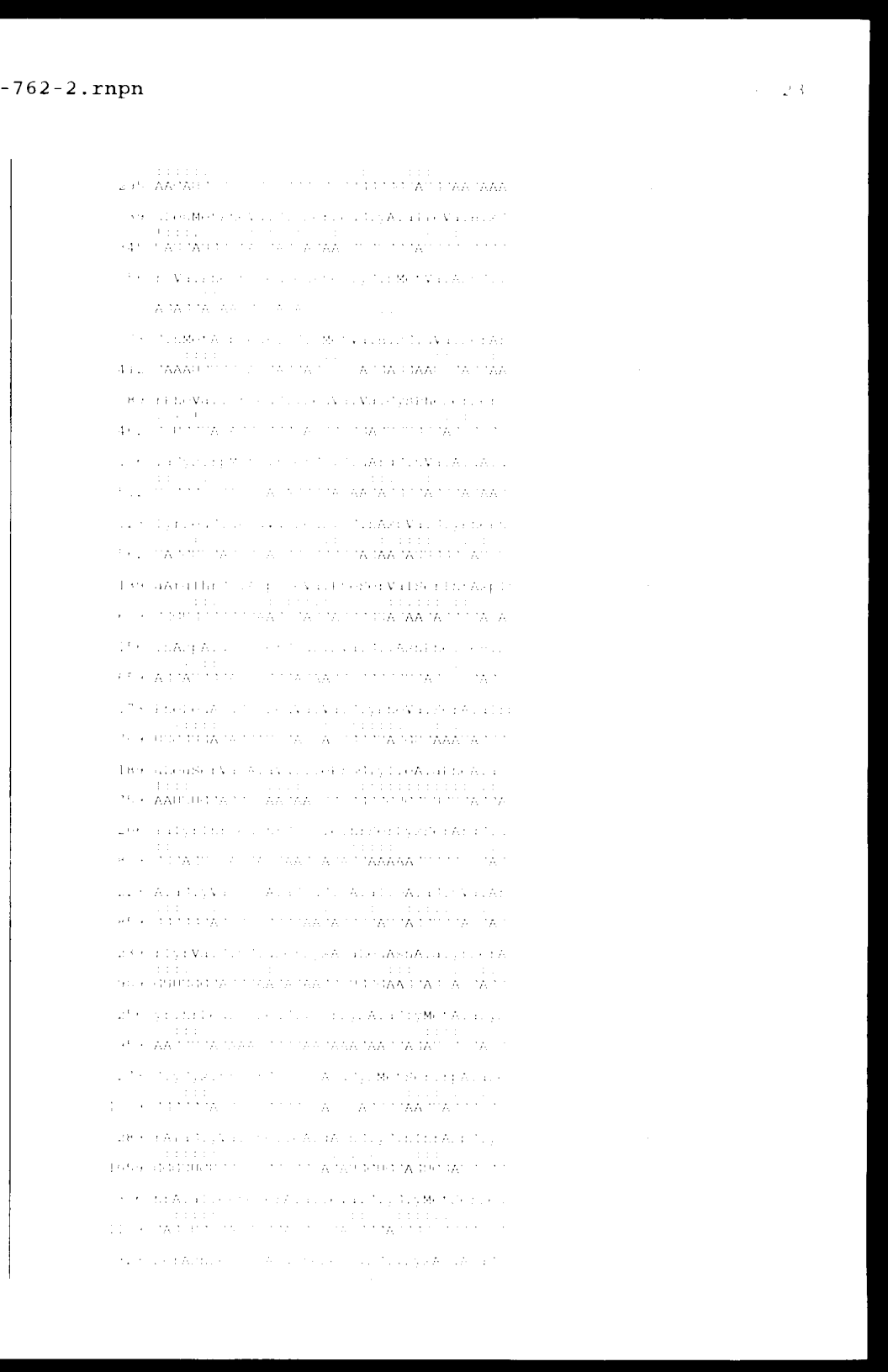


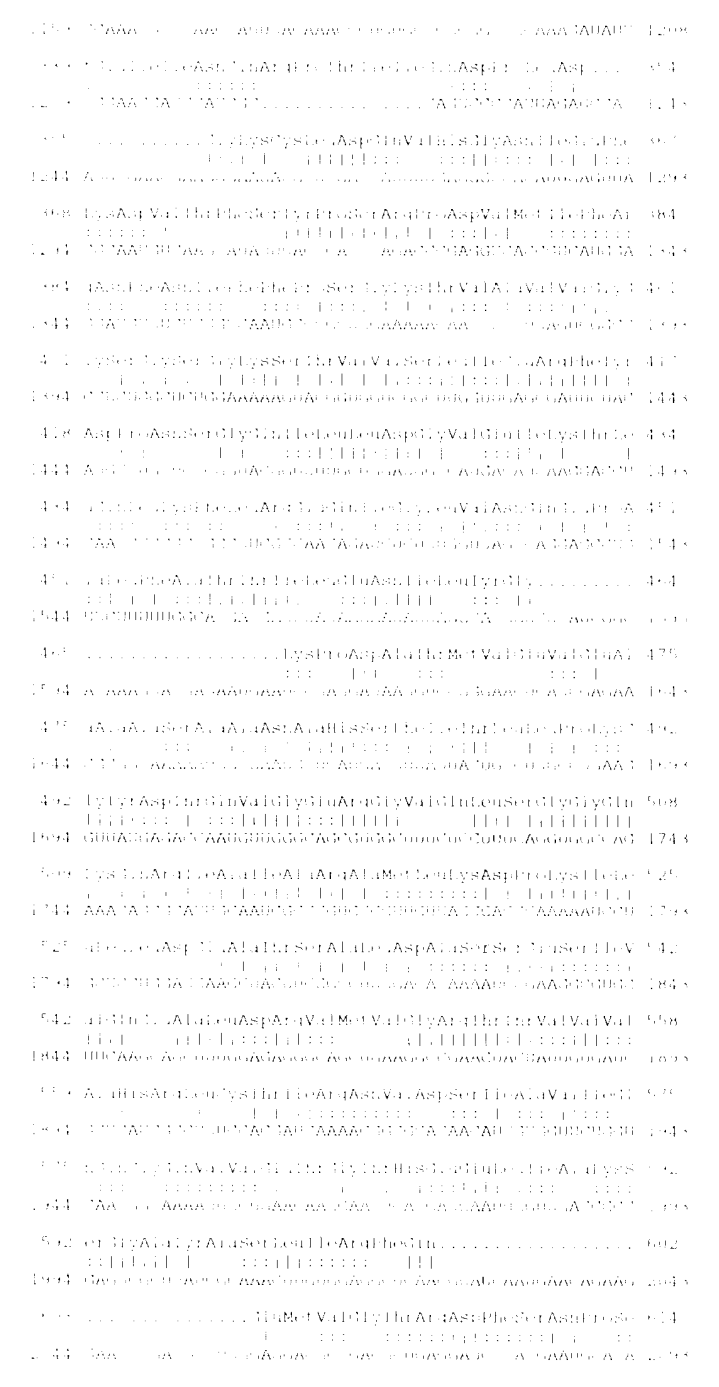


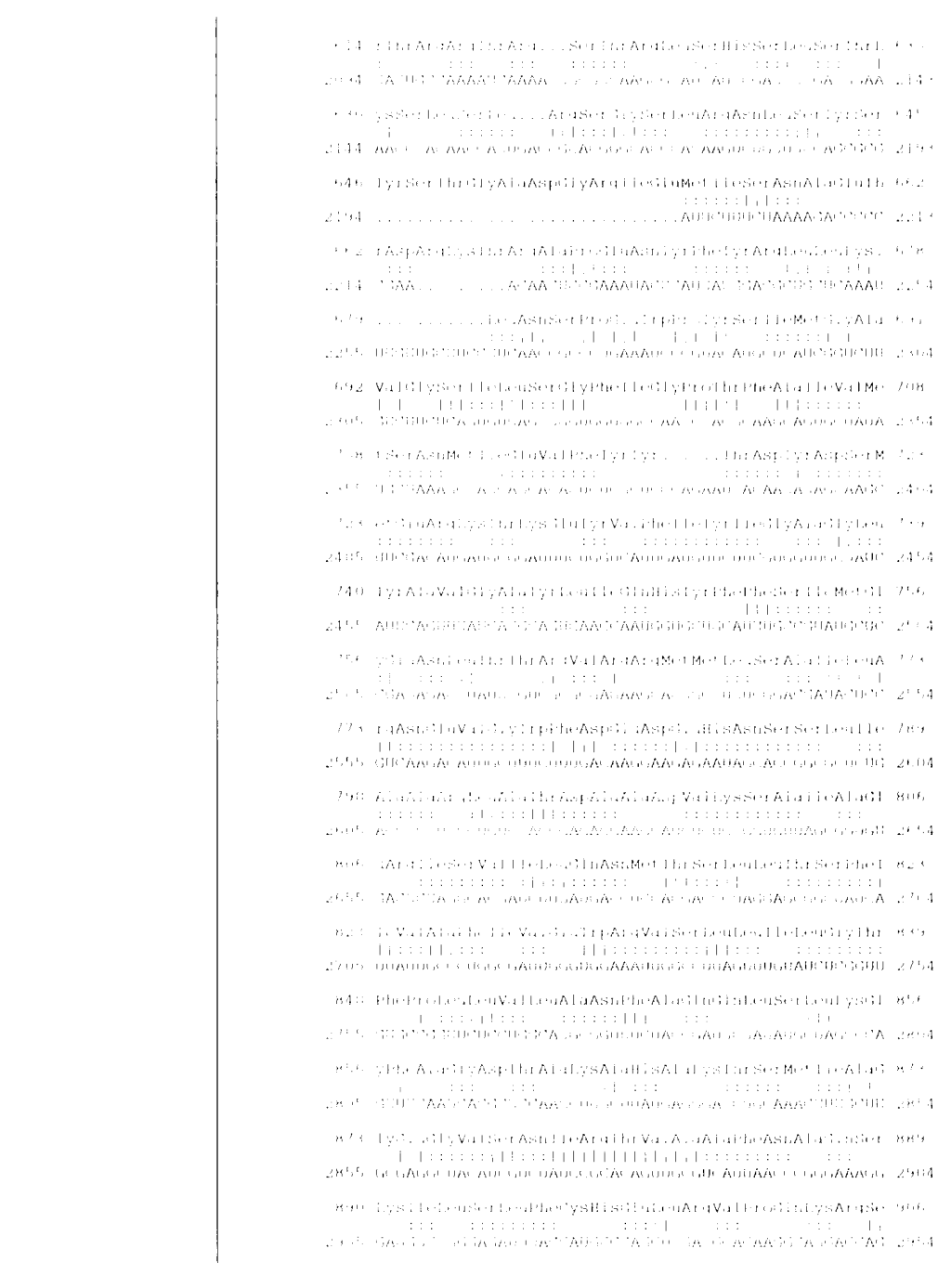


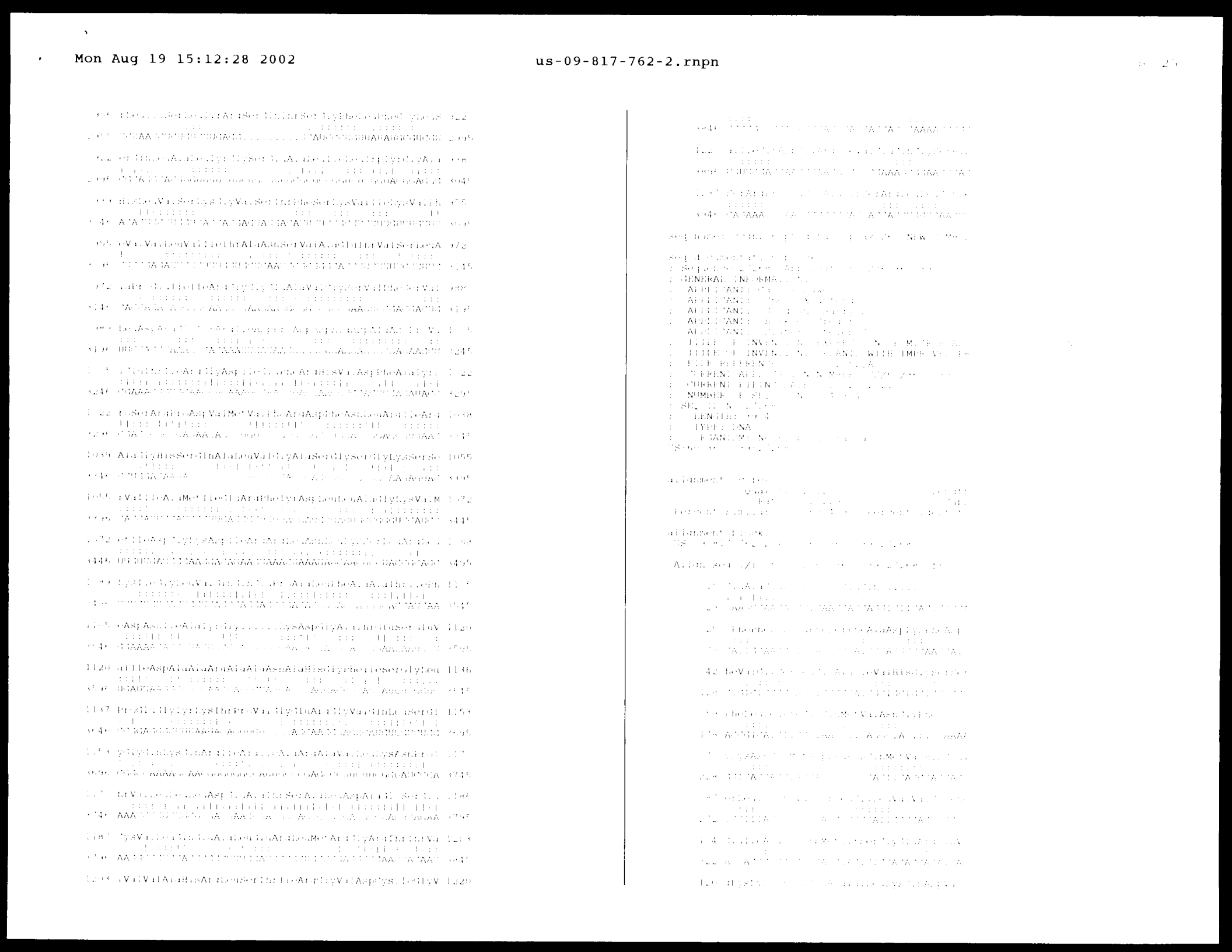


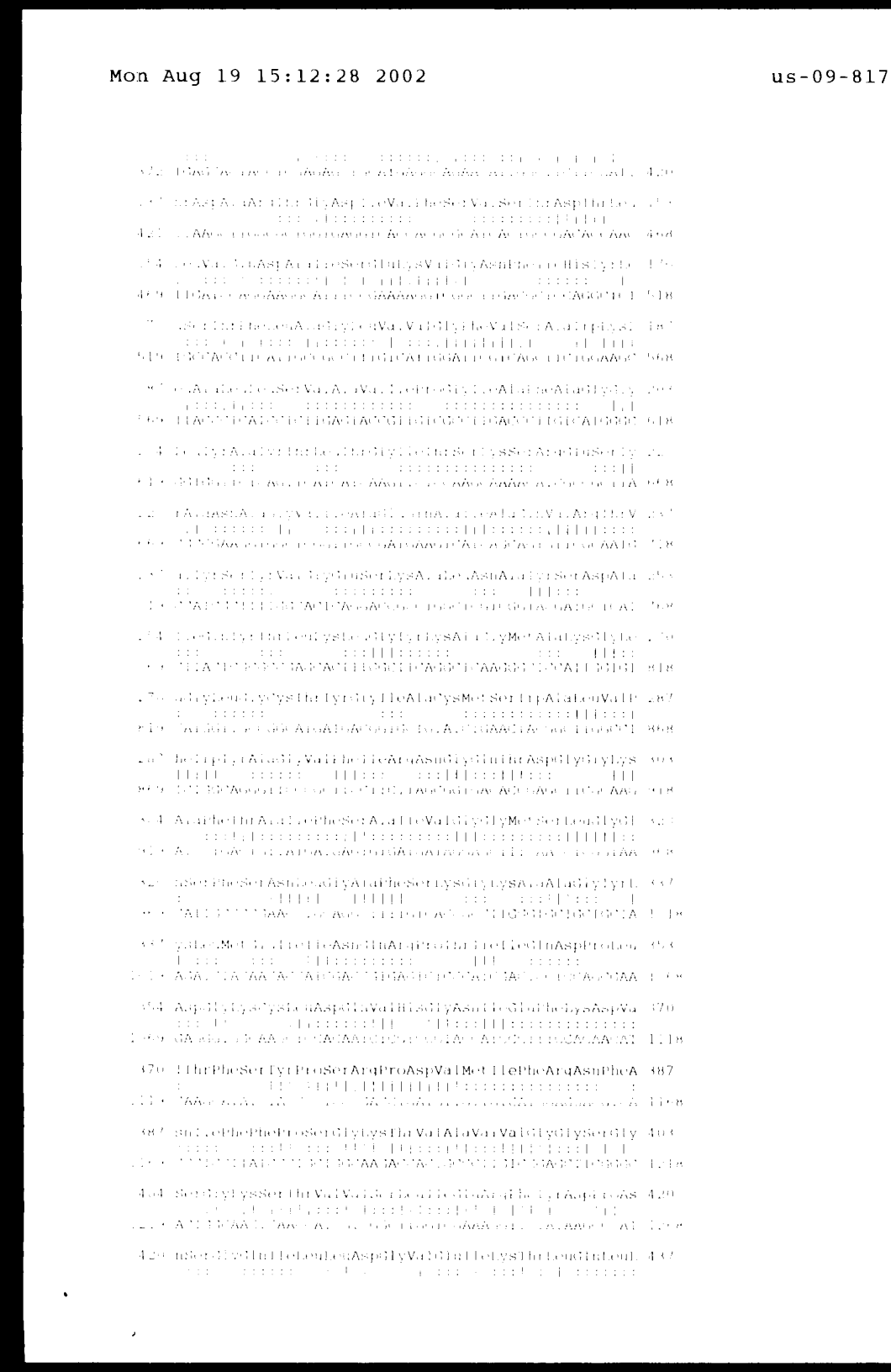


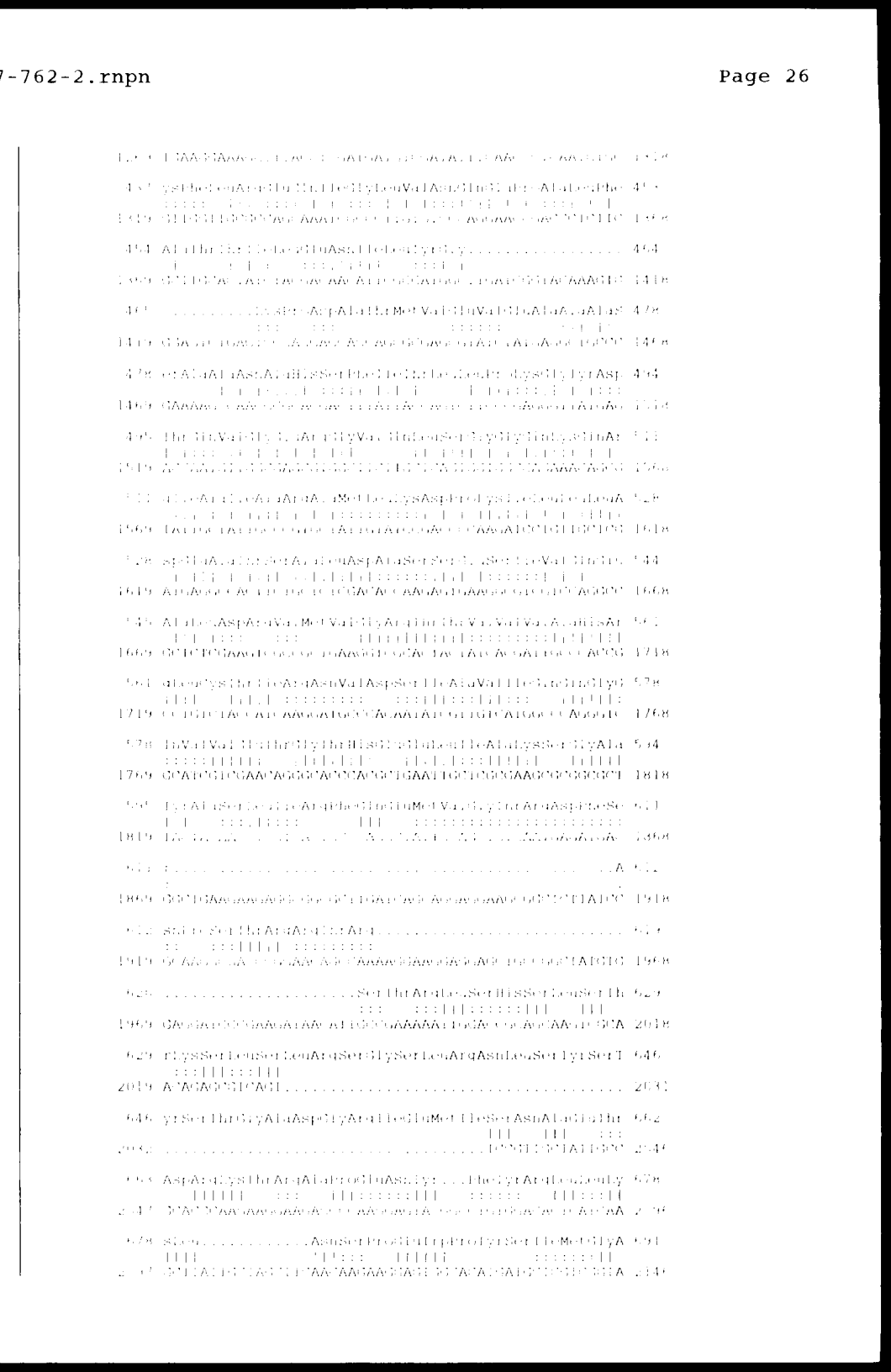


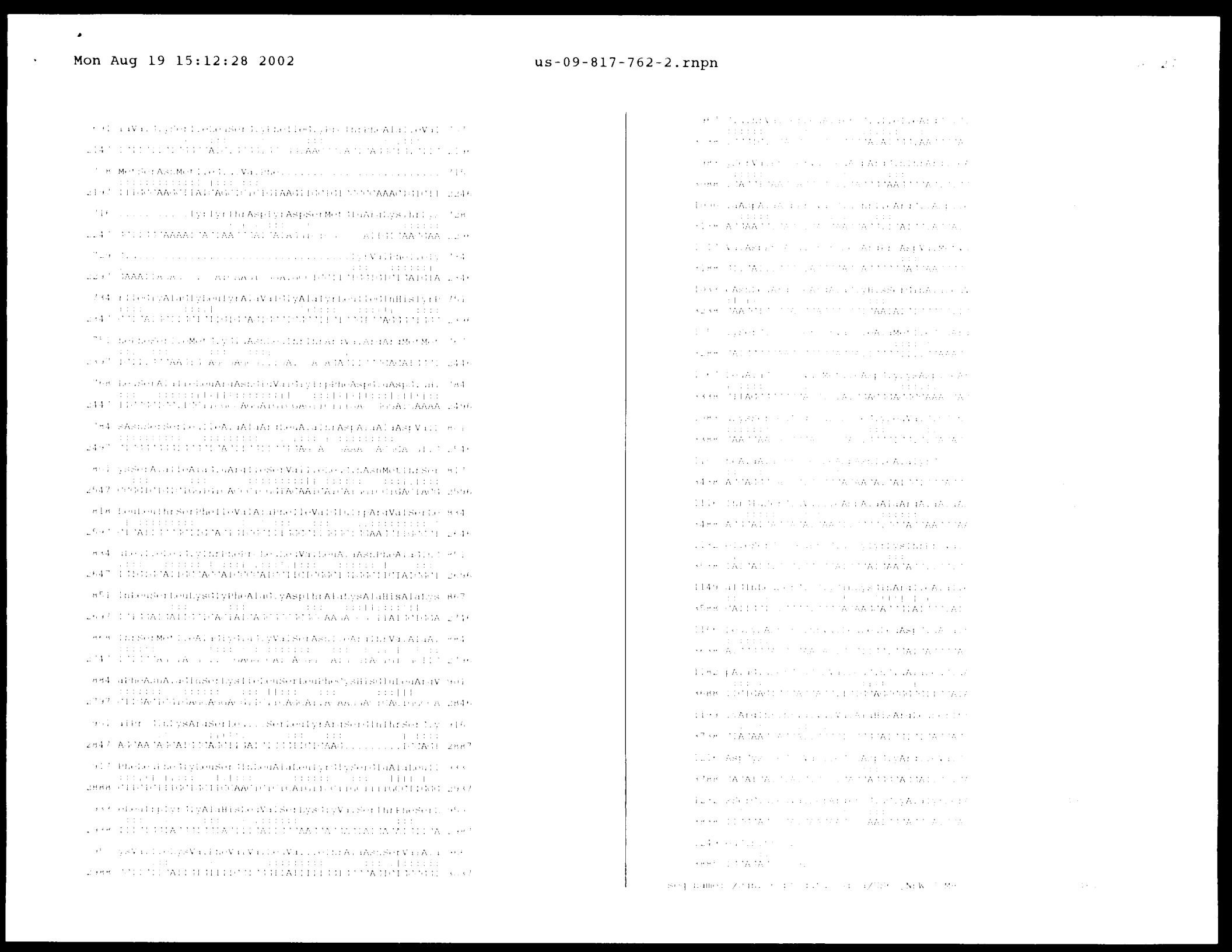


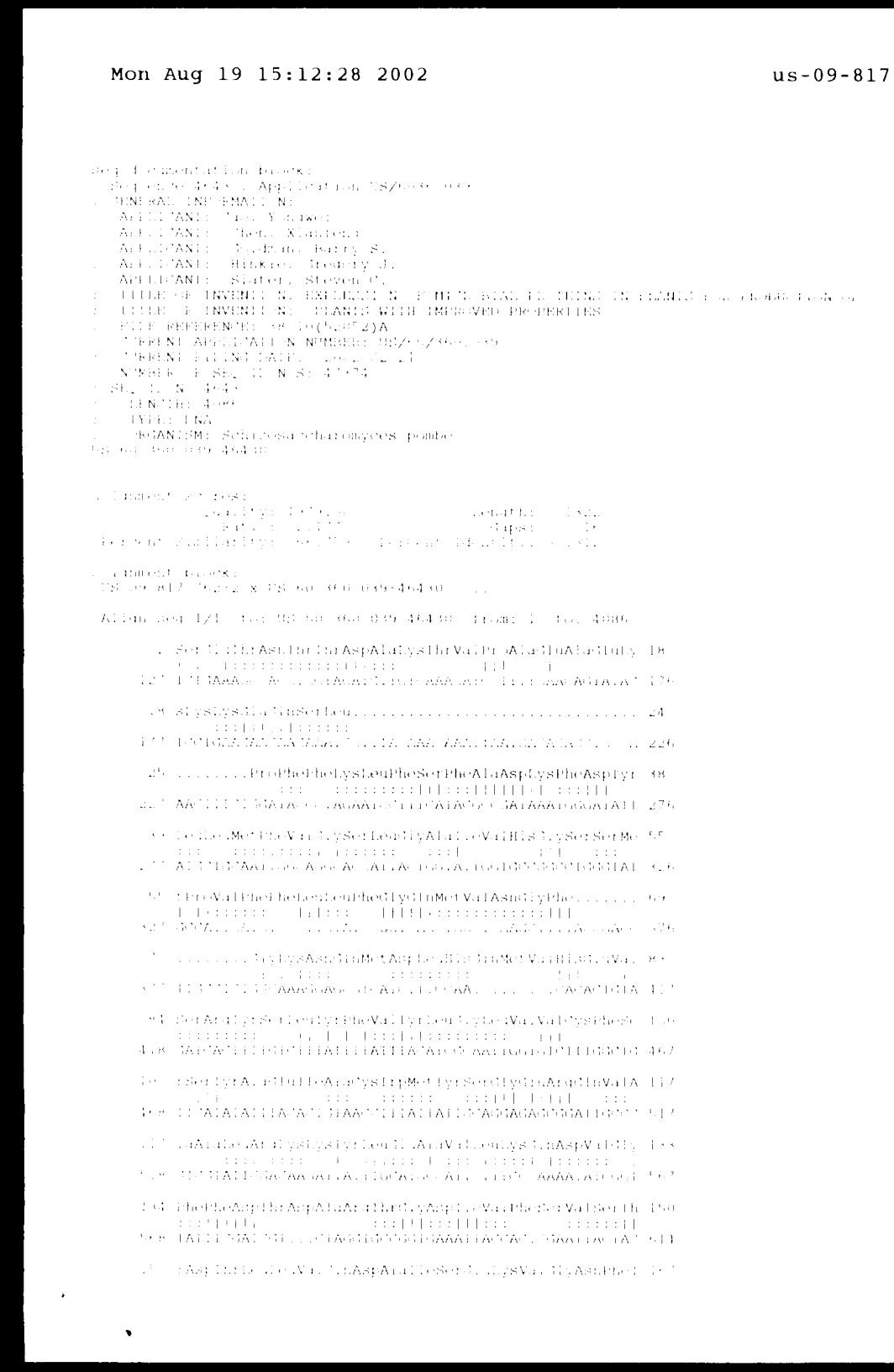


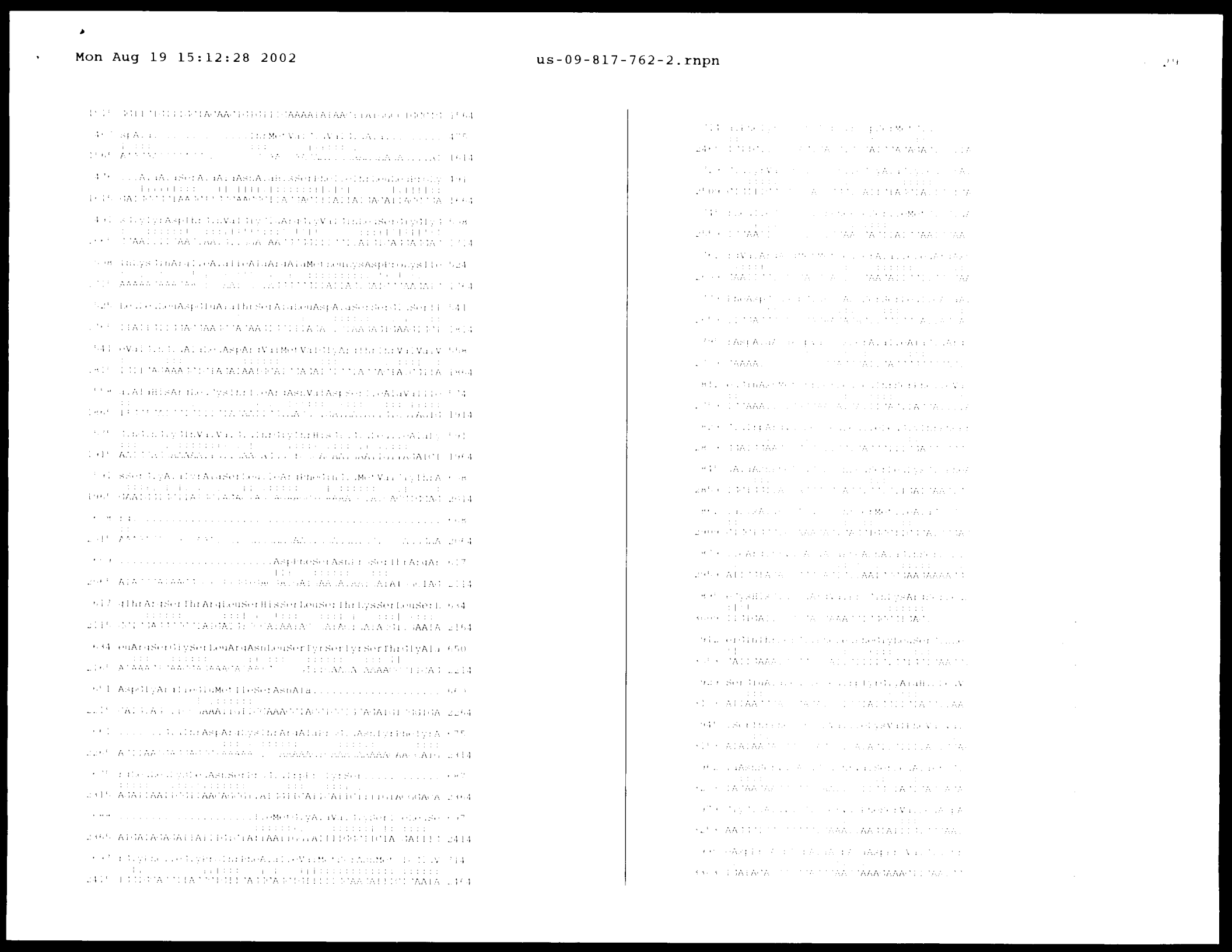


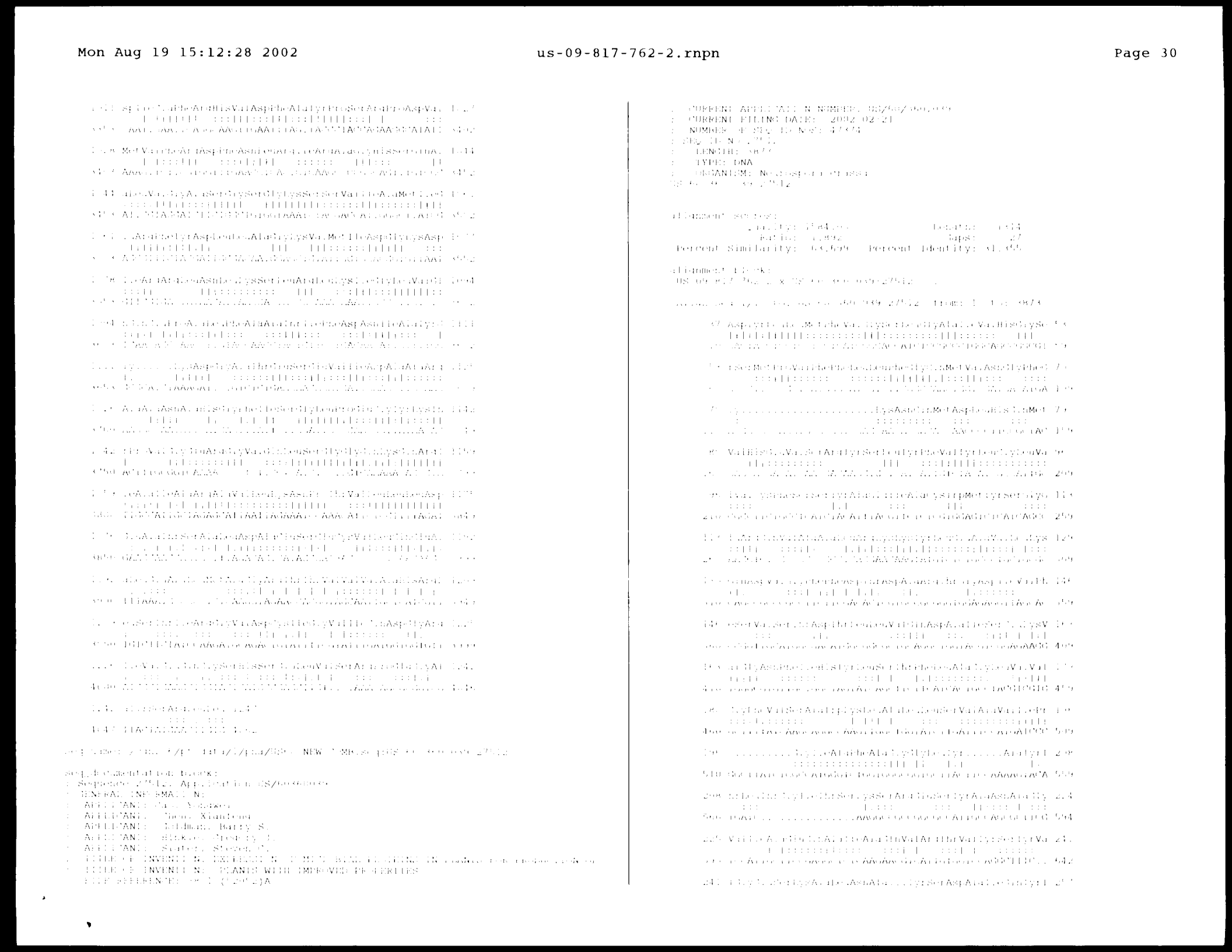


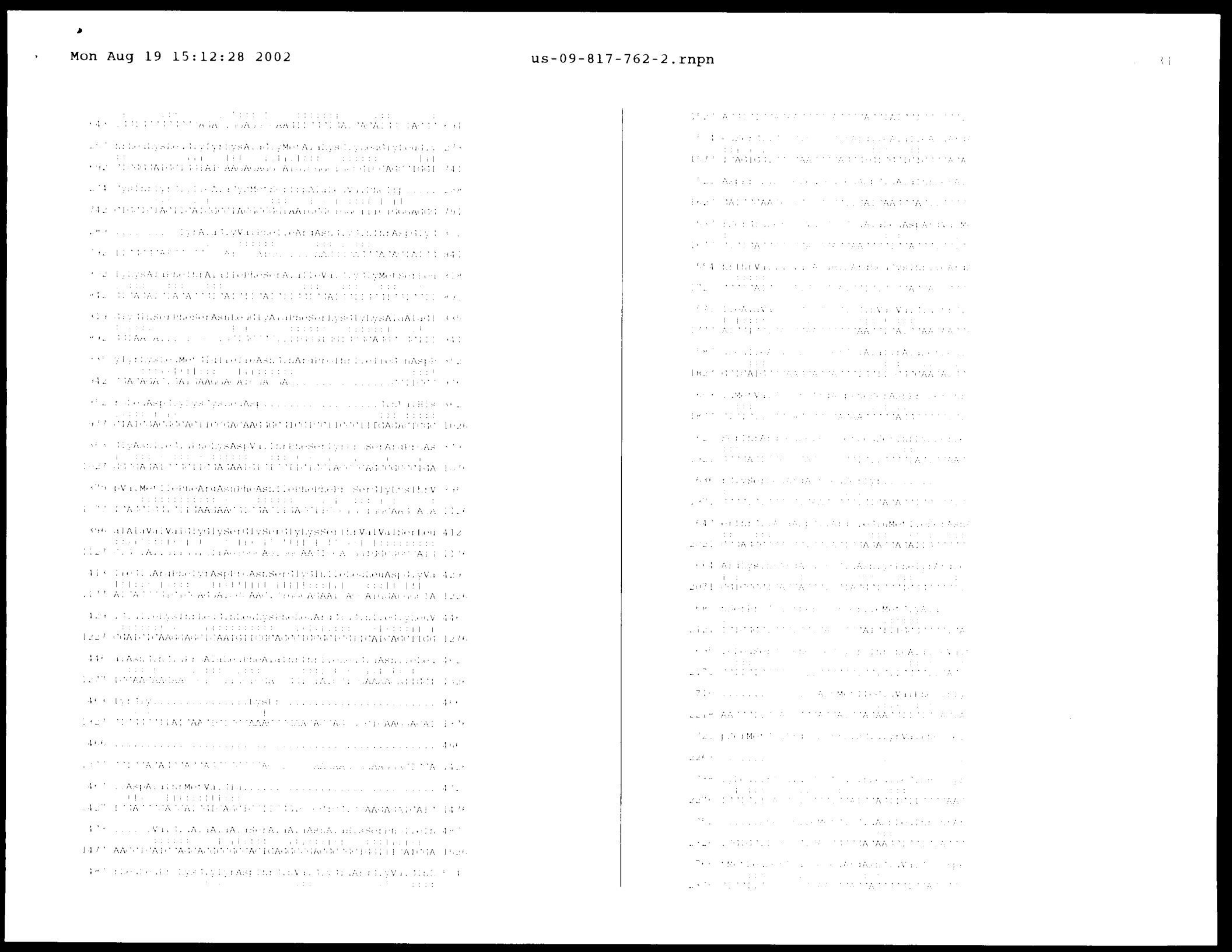


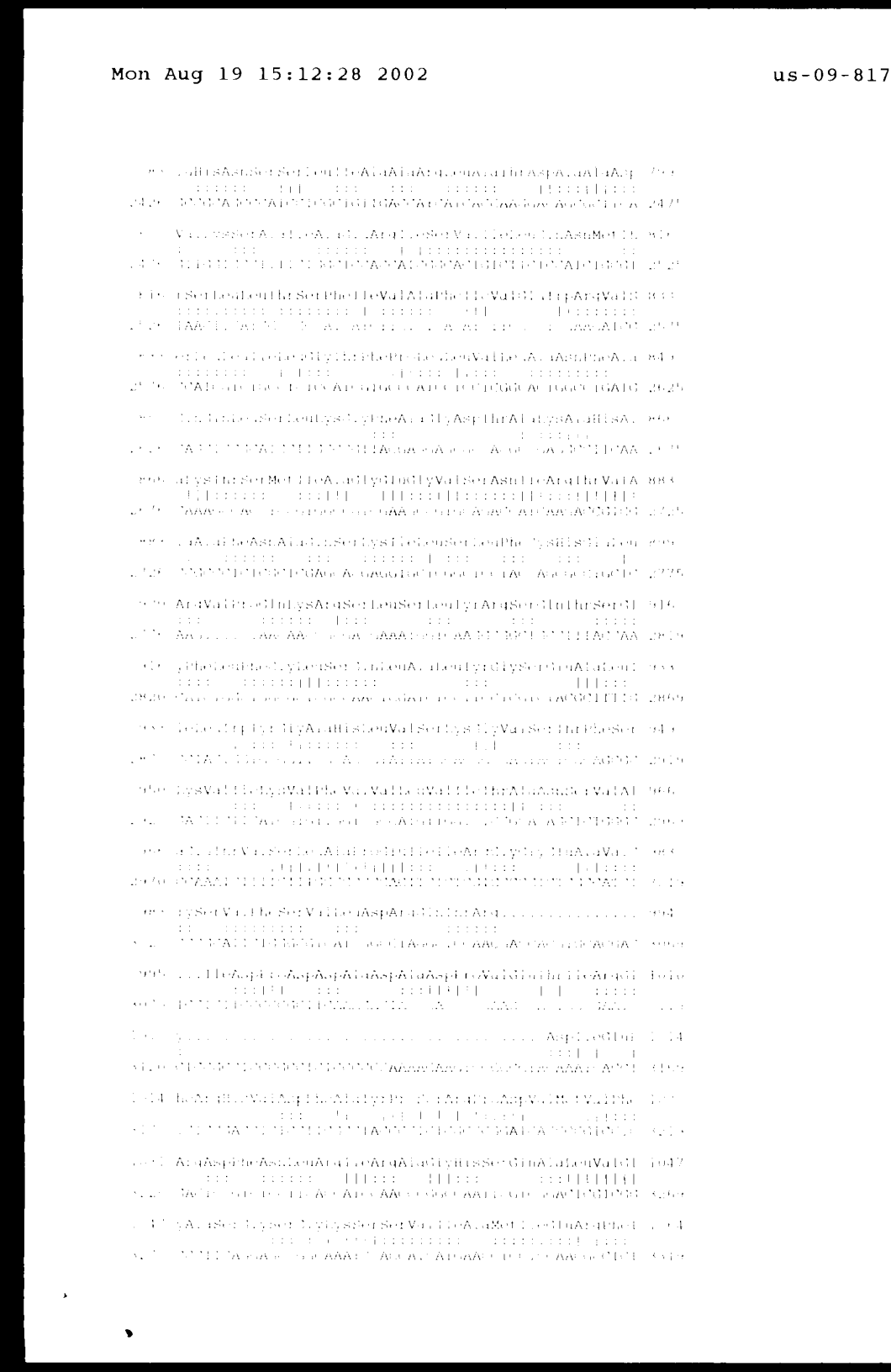


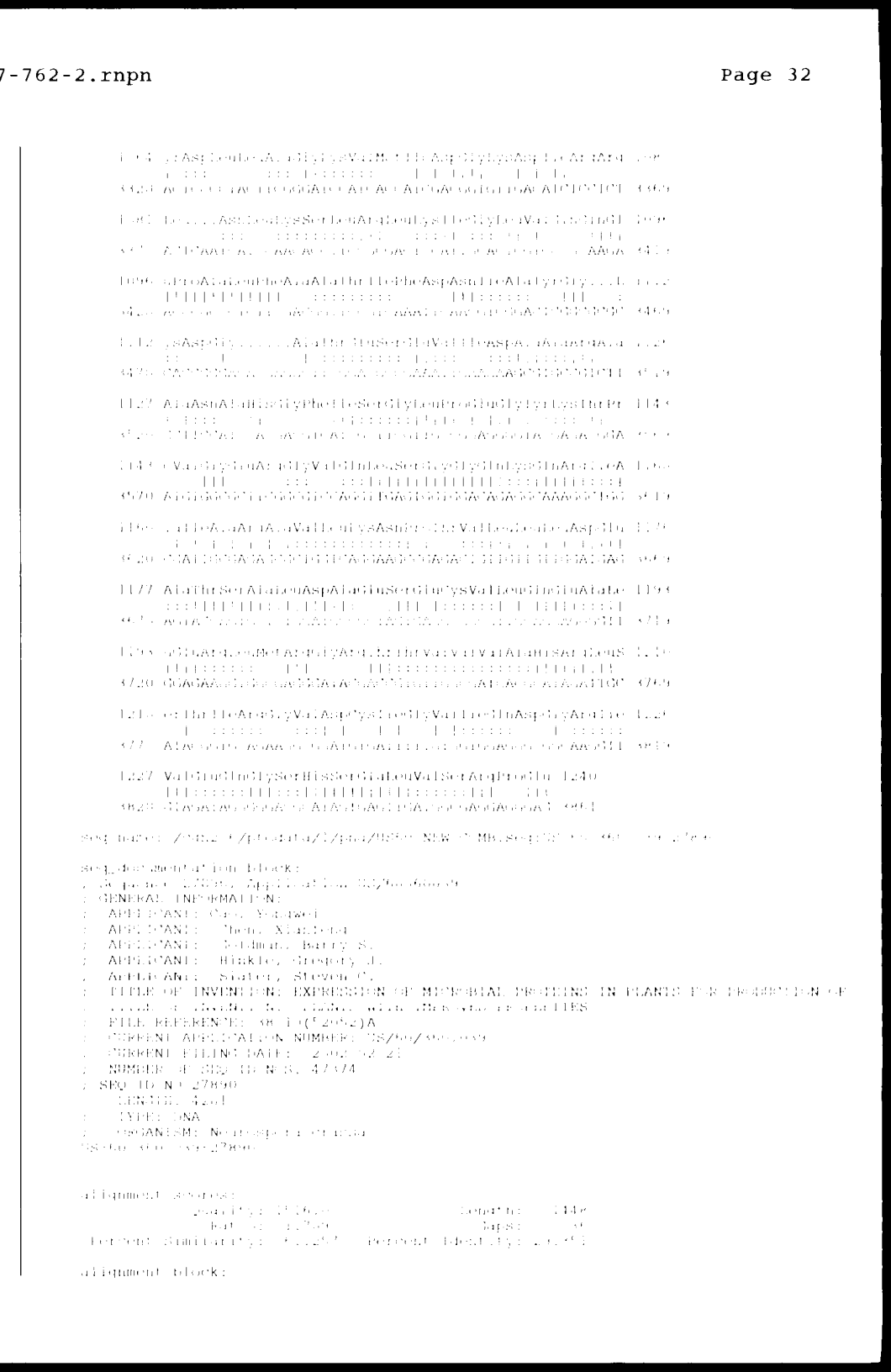


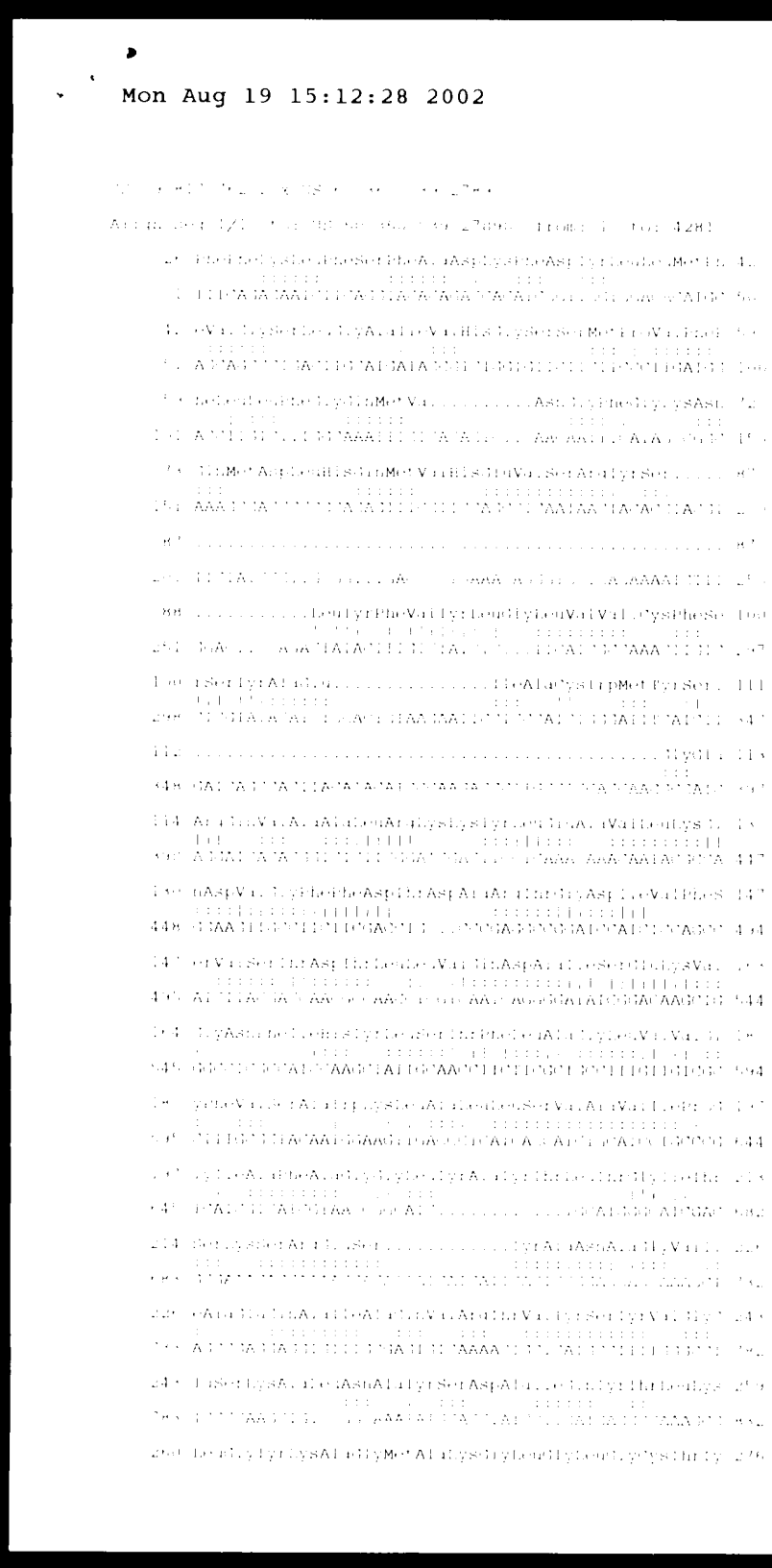


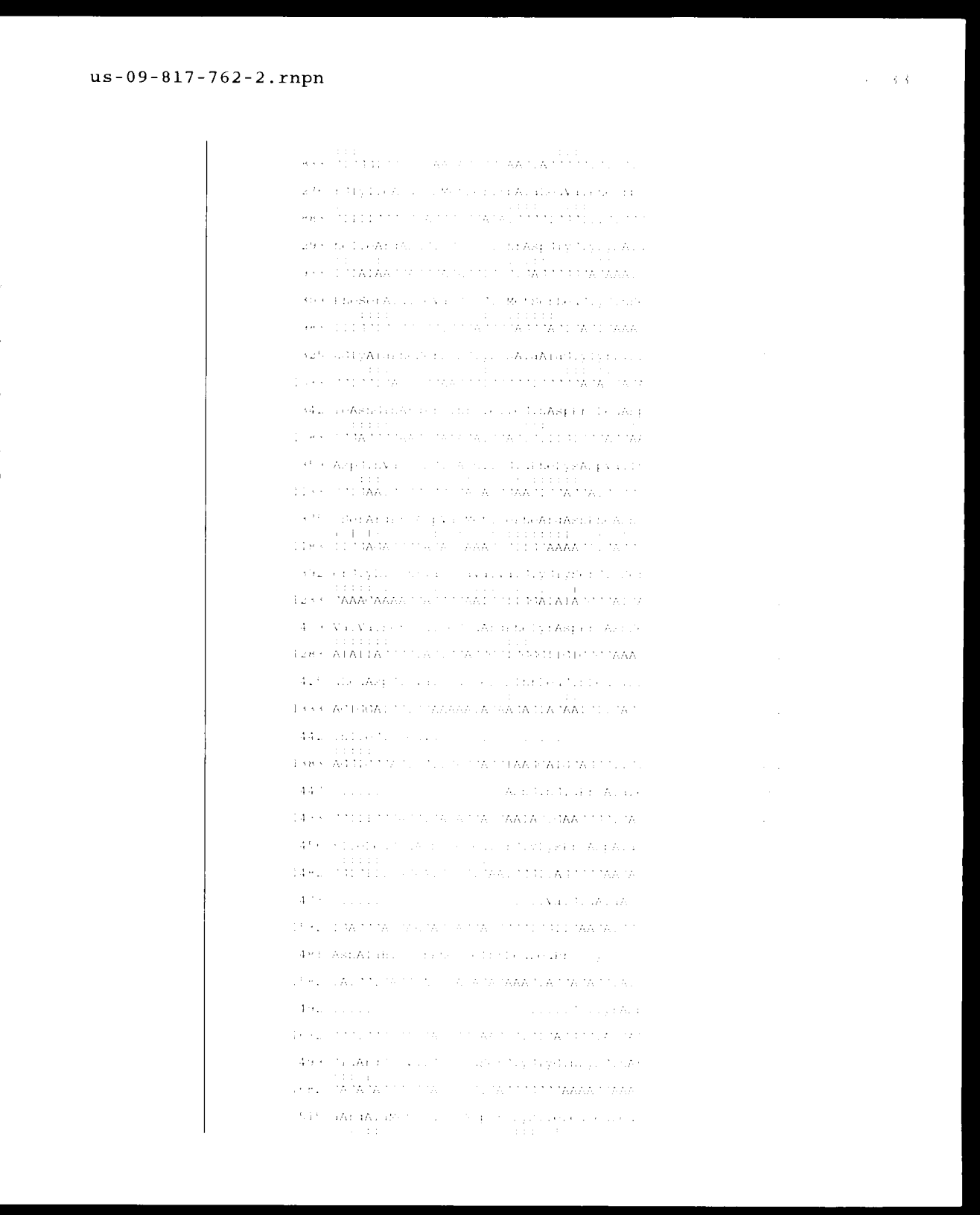


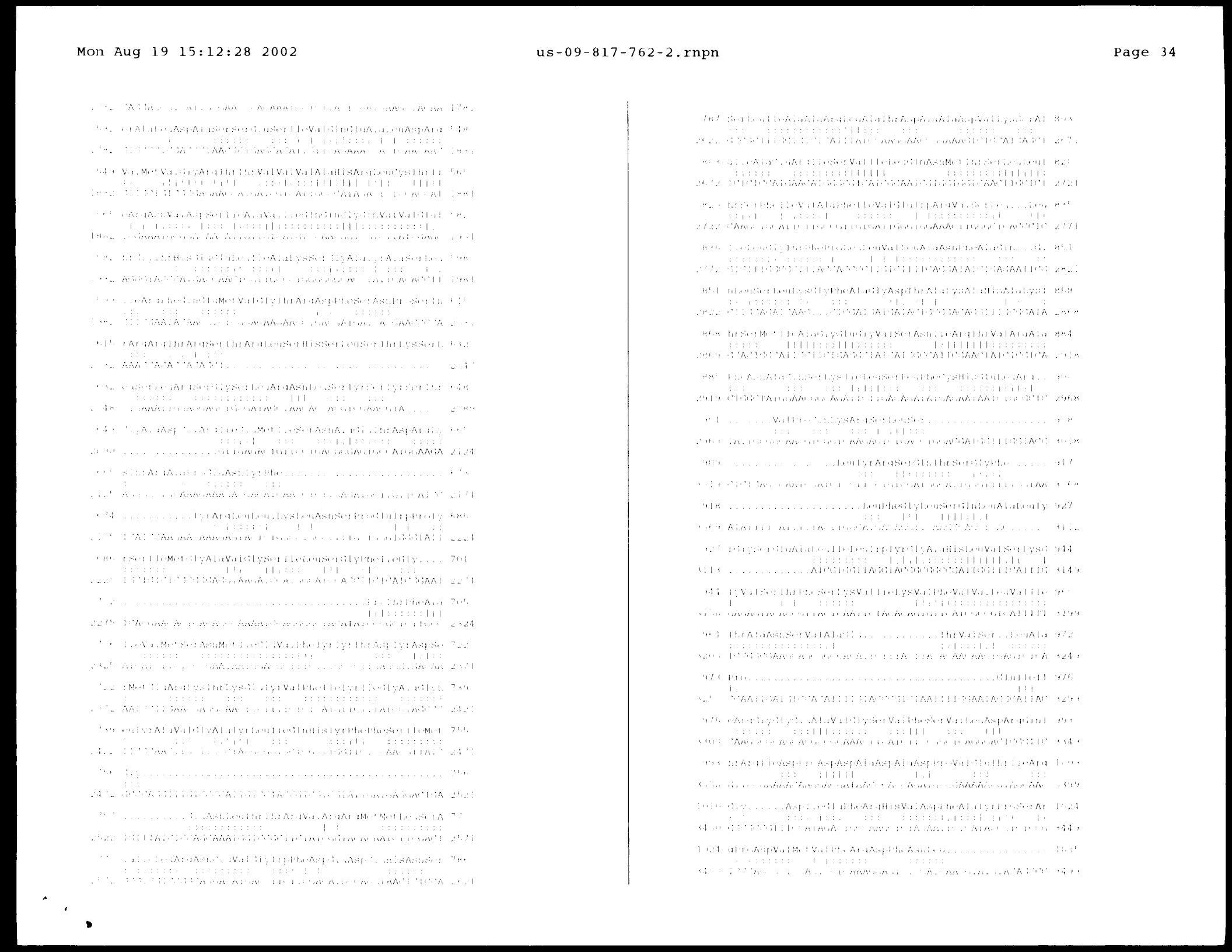


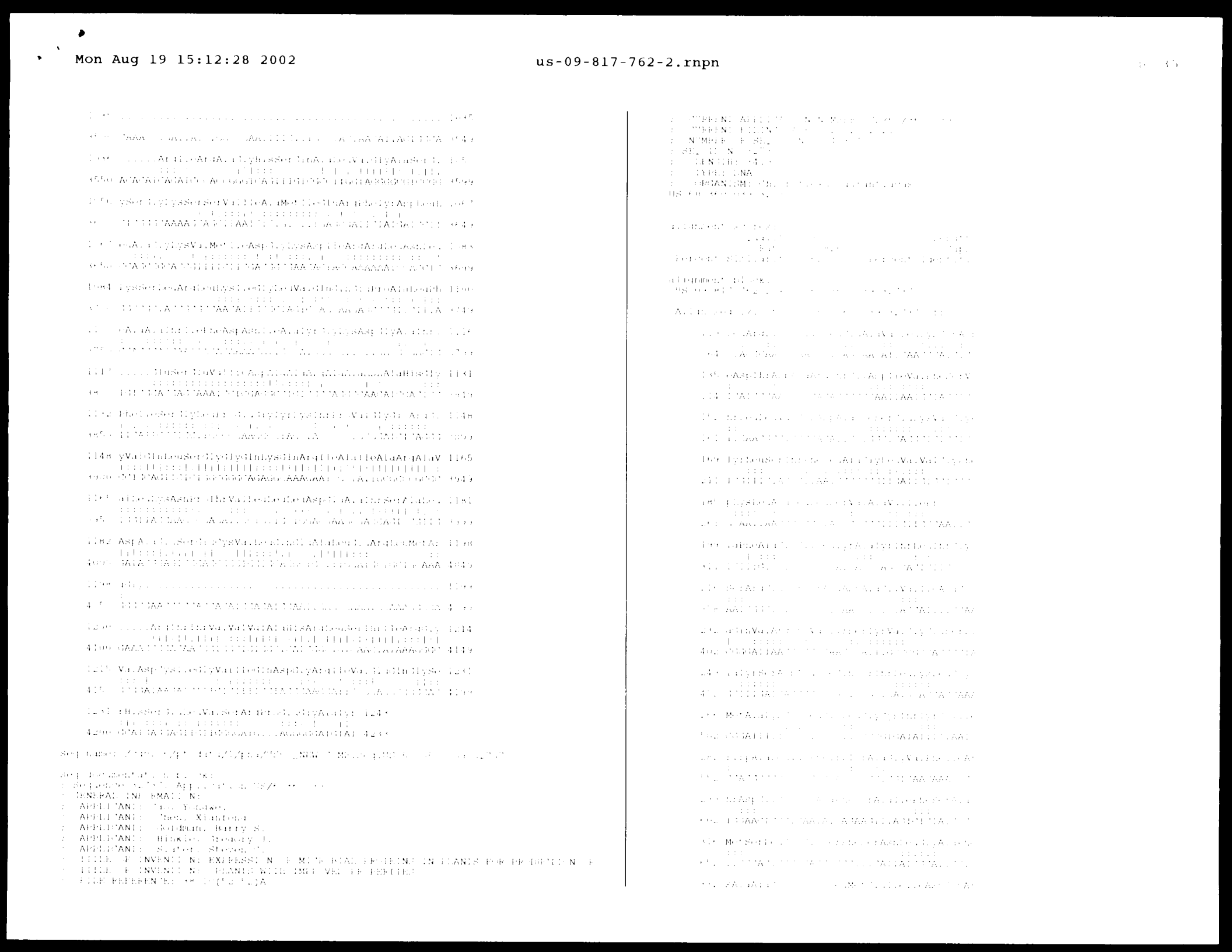


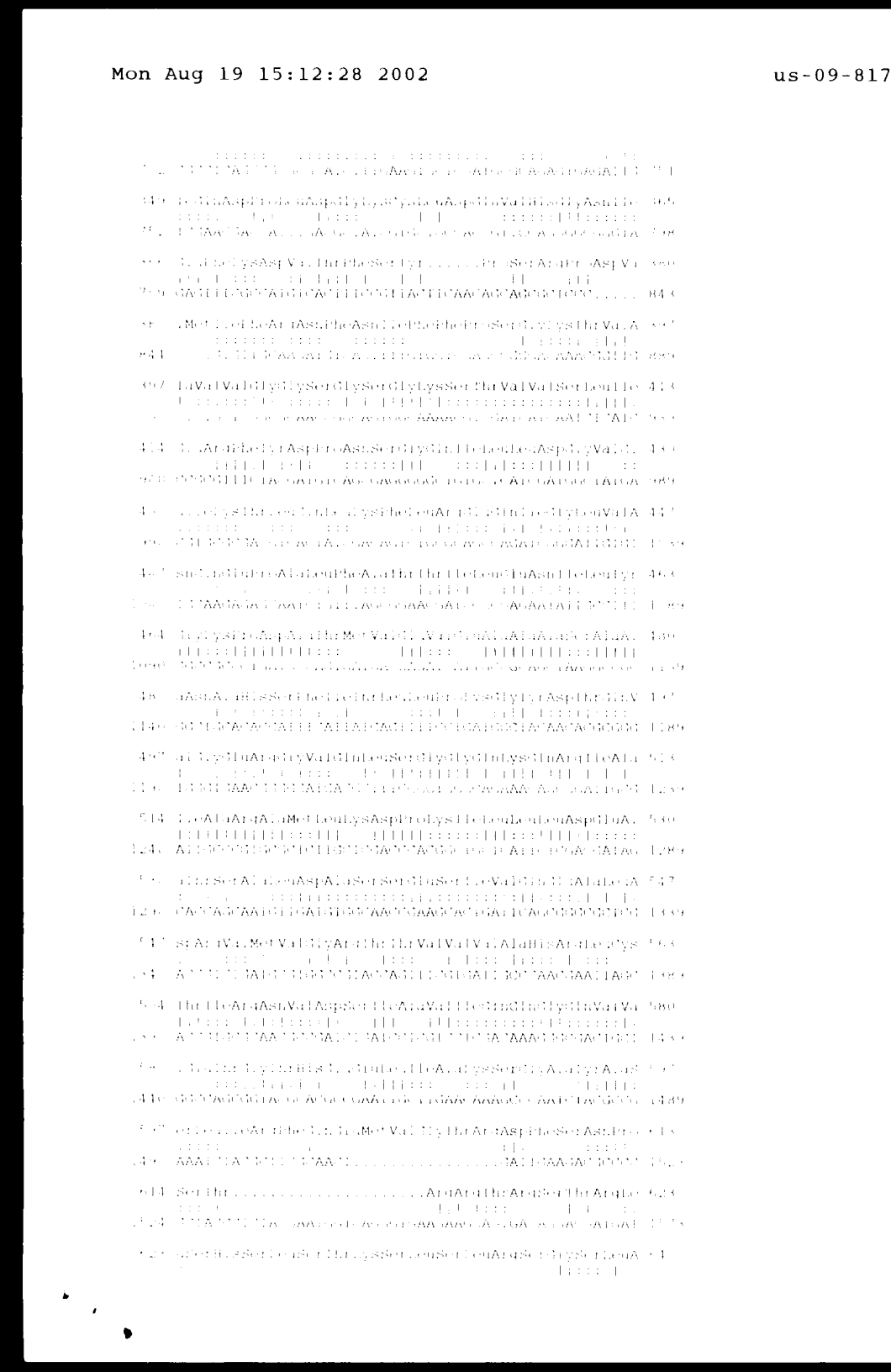


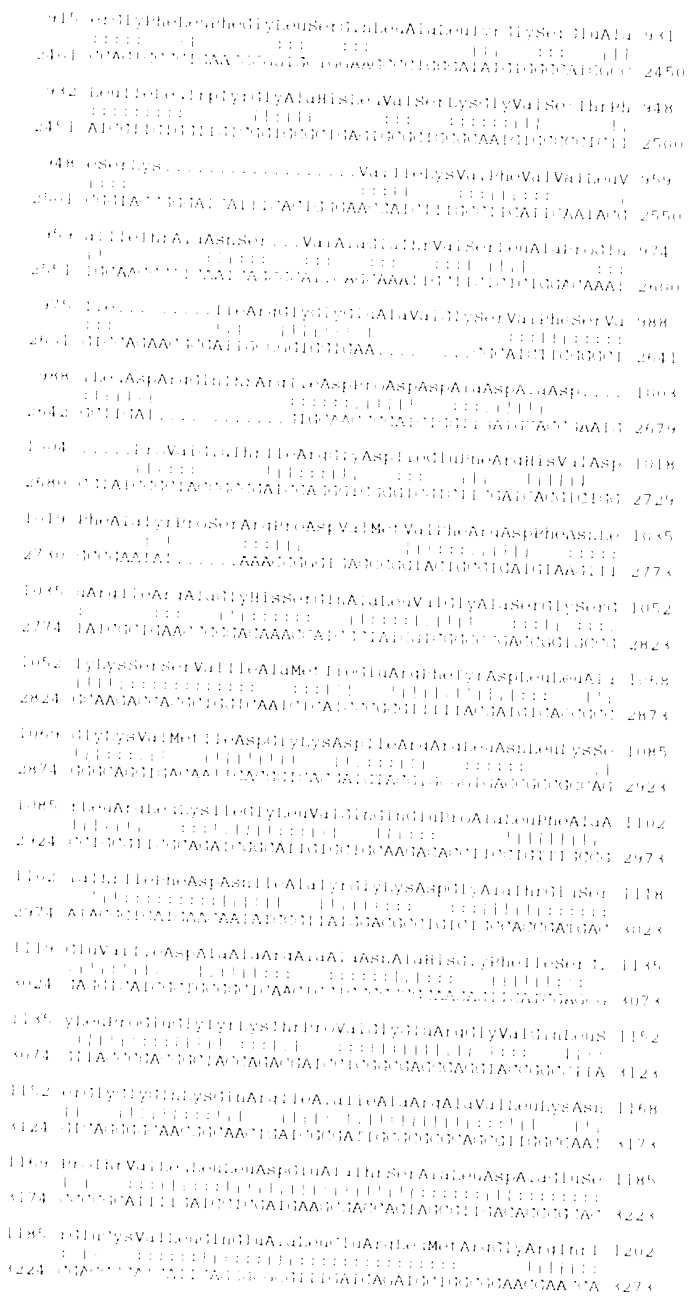


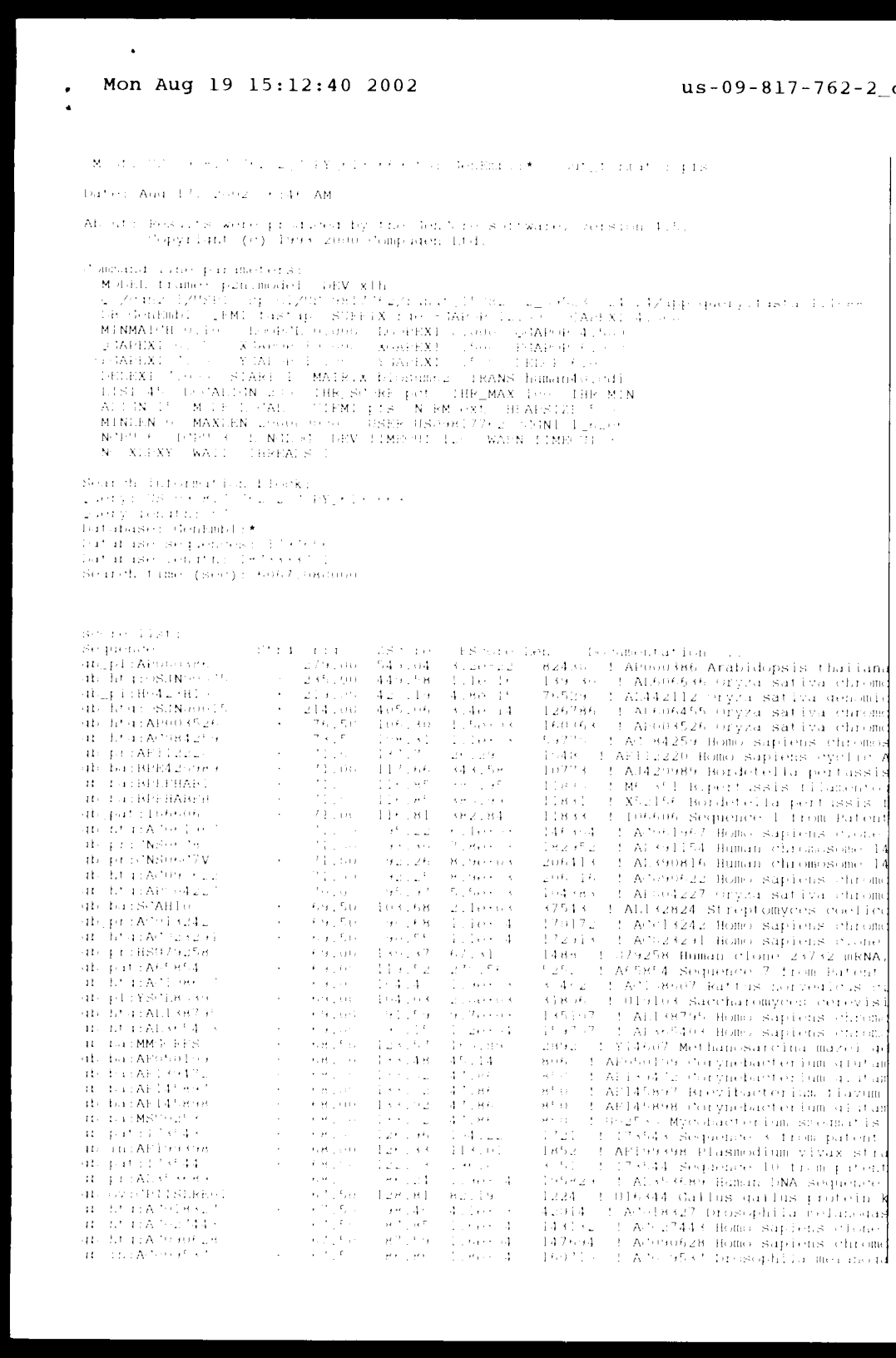


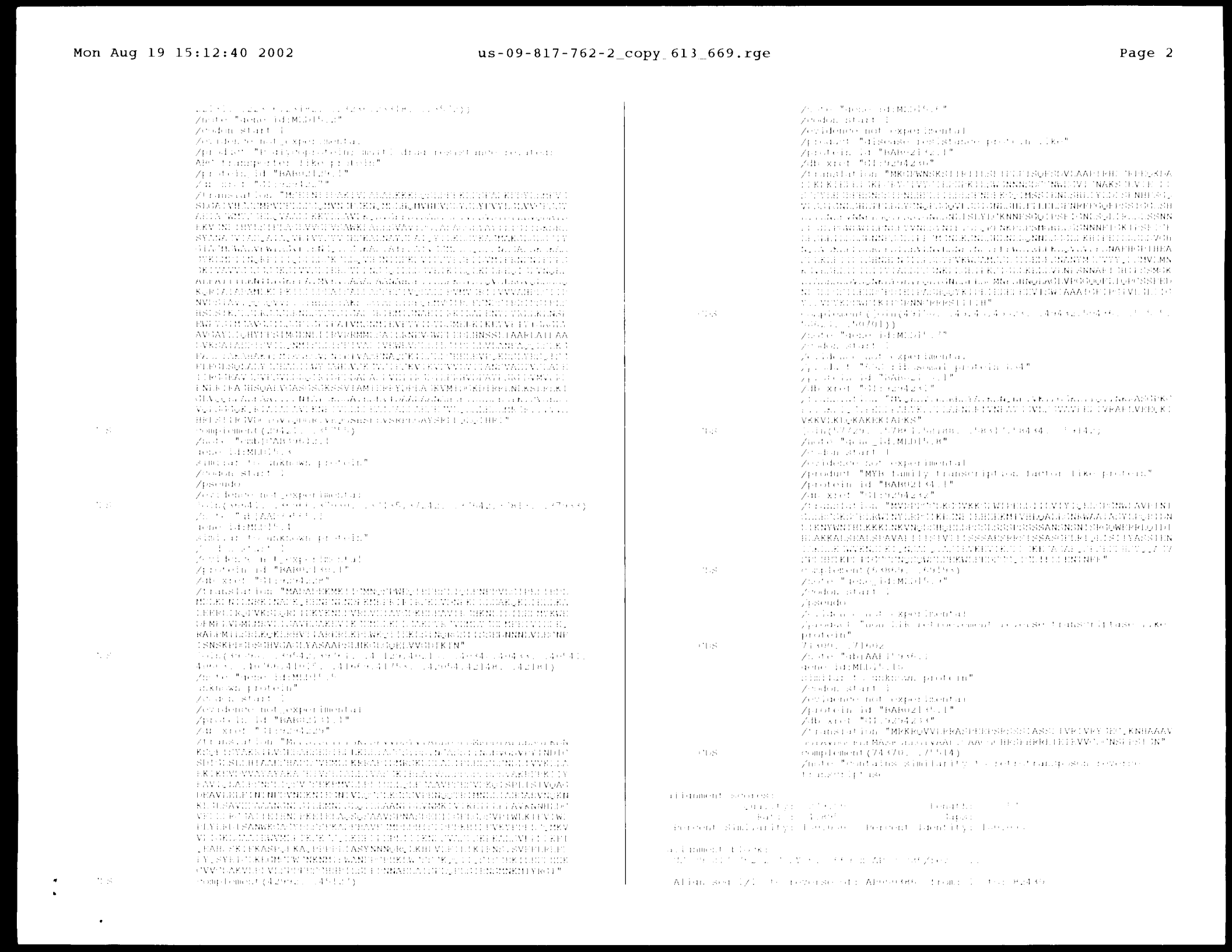


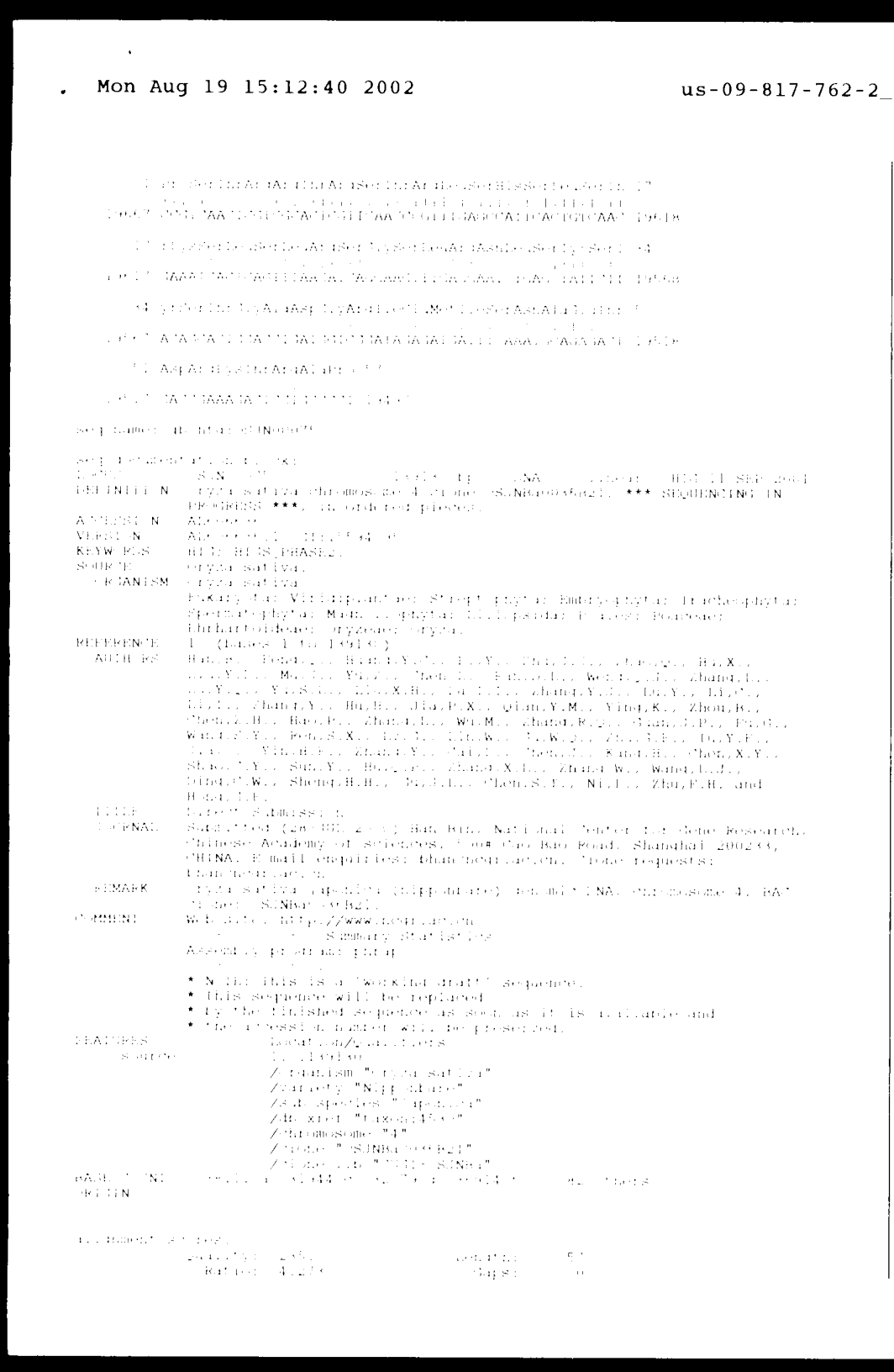


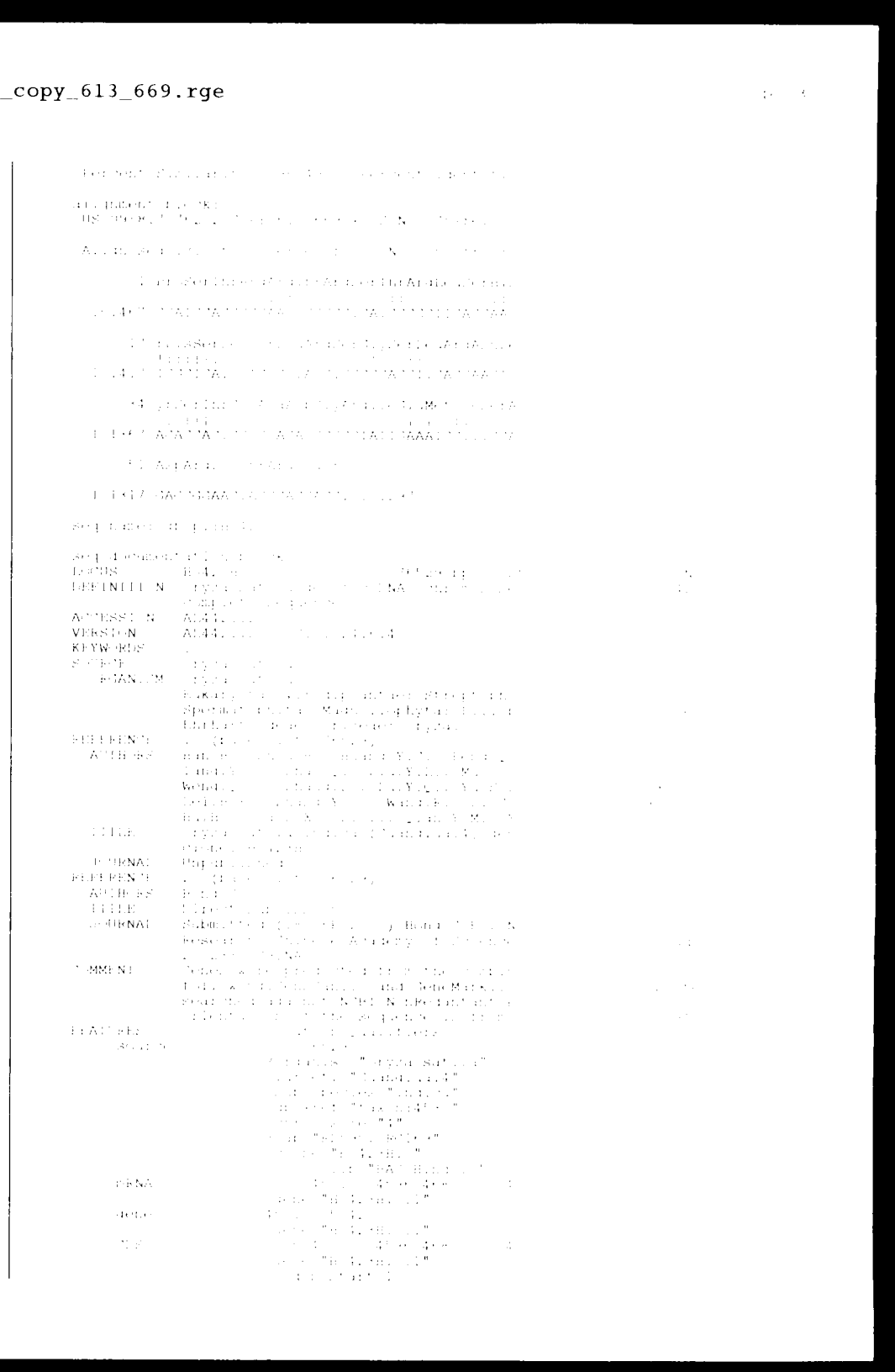




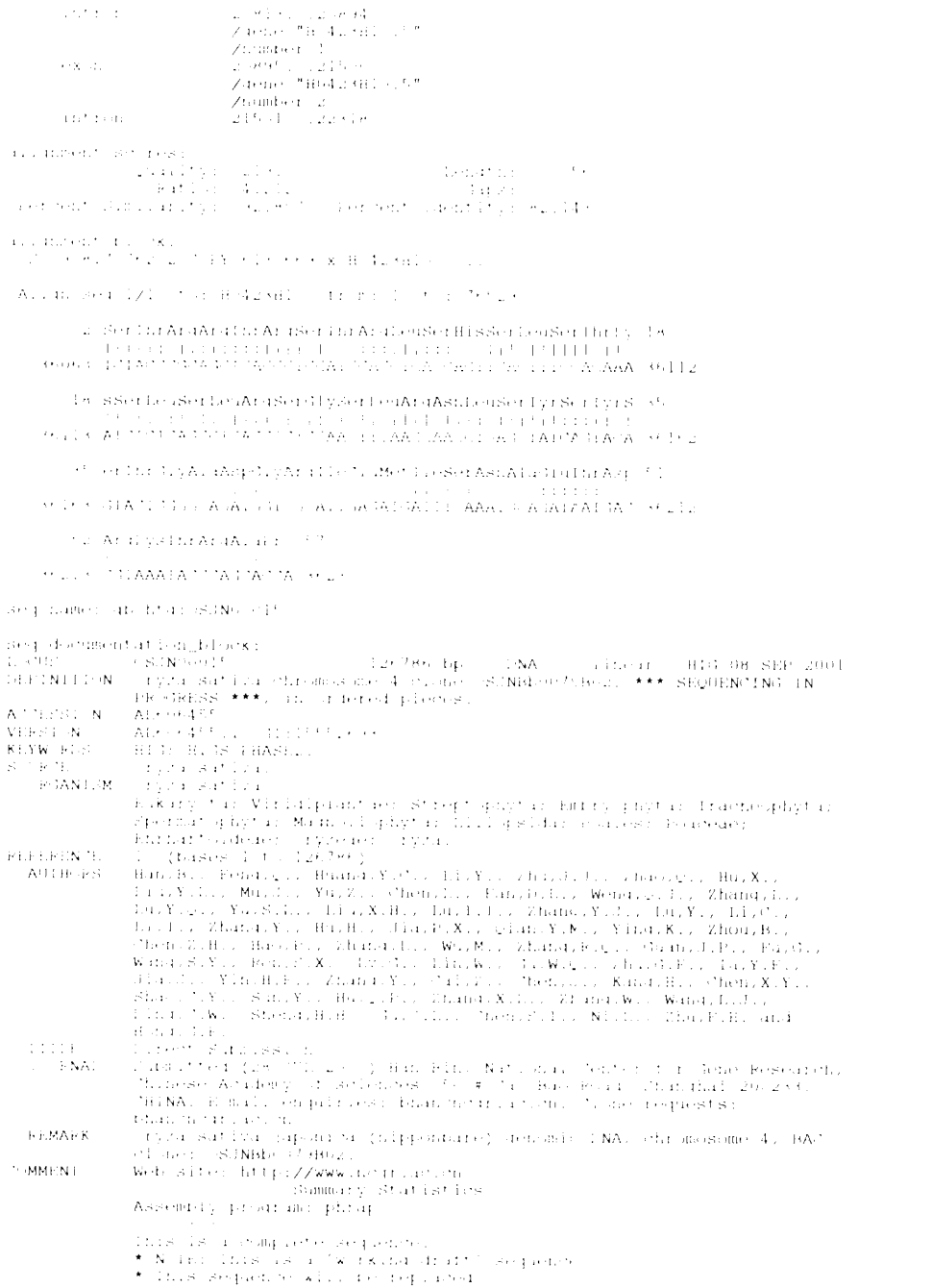






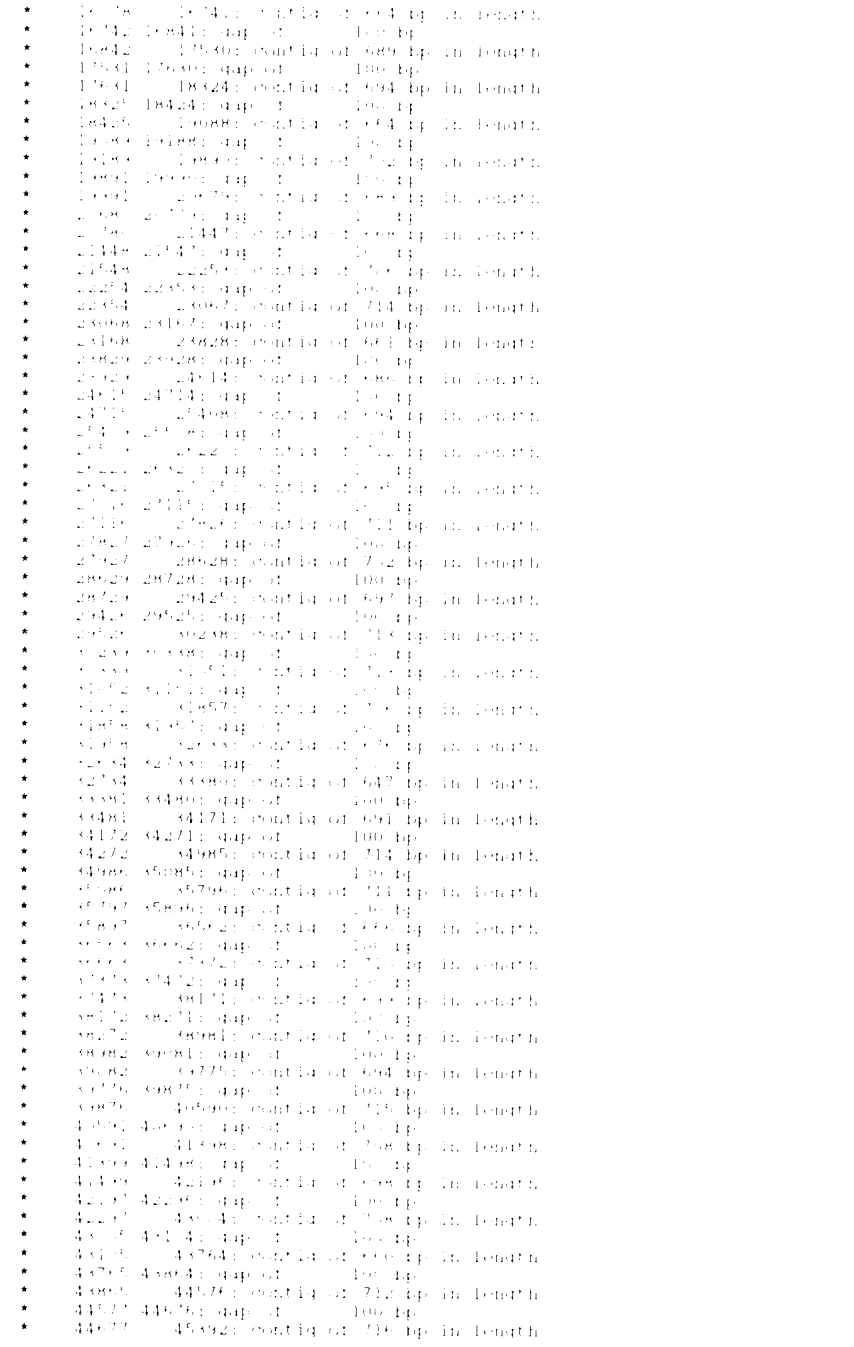


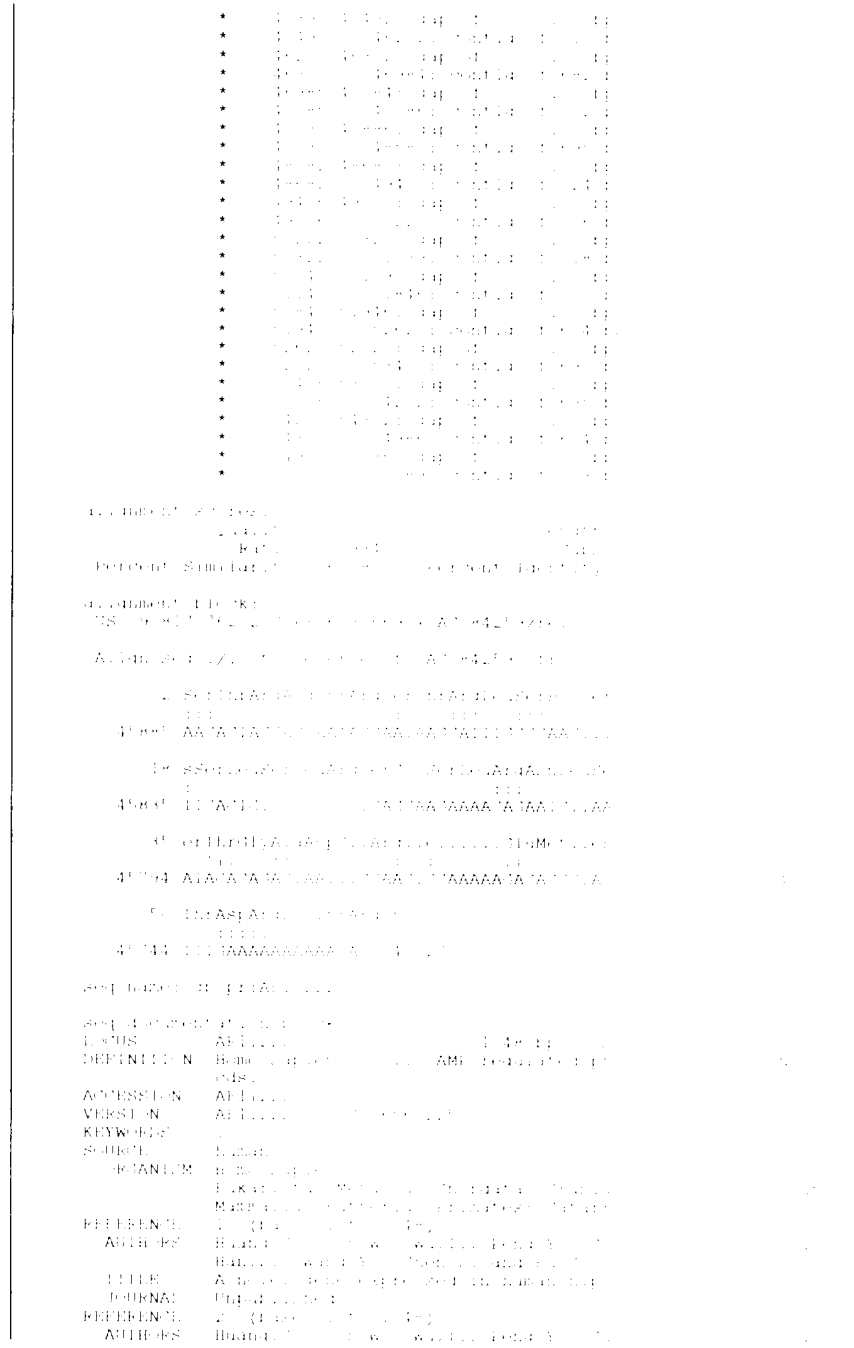
```
                     /product "hypothetical protein"
                     /protein_id "CAC09458.1"
                     /db_xref "GI:10241618"
                     /translation "MEEWKEIIKKSSAAIAFTTAAKSGKELGKGYLLKLIKTGKYSGTE
                     LNKEIP"
     misc_feature    8888..9775
                     /gene "H042H10.4"
                     /number 1
     CDS             9888..9775
                     /gene "H042H10.4"
                     /number 2
     misc_feature    9775..9229
                     /gene "H042H10.4"
                     /number 3
     exon            12170..12451
                     /gene "H042H10.4"
                     /number 4
     misc_feature    12552..12414
                     /gene "H042H10.4"
                     /number 5
     exon            12342..12567
                     /gene "H042H10.4"
                     /number 6
     misc_feature    12556..12714
                     /gene "H042H10.4"
                     /number 7
     exon            12746..13041
                     /gene "H042H10.4"
                     /number 8
     misc_feature    13042..13414
                     /gene "H042H10.4"
                     /number 9
     exon            13416..13570
                     /gene "H042H10.4"
                     /number 10
     exon            complement(13442..13446)
                     /gene "H042H10.4"
                     /number 11
     mRNA            complement(join(13442..13446,16976..17600))
                     /gene "H042H10.4"
     gene            complement(13442..17600)
                     /gene "H042H10.4"
     CDS             complement(join(13442..13446,16976..17600))
                     /gene "H042H10.4"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09458.1"
                     /db_xref "GI:10241618"
                     /translation "MEEWKEIIKKSSAAIAFTTAAKSGKELGKGYLLKLIKTGKYSGTE
                     LNKEIP"
     misc_feature    complement(15647..16972)
                     /gene "H042H10.4"
                     /number 12
     exon            complement(16976..17600)
                     /gene "H042H10.4"
                     /number 13
     mRNA            join(20660..20812,20895..21000,22419..22634)
                     /gene "H042H10.5"
     gene            20660..22634
                     /gene "H042H10.5"
     CDS             join(20660..20812,20895..21000,22419..22634)
                     /gene "H042H10.5"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC09457.1"
                     /db_xref "GI:10241617"
                     /translation "MGMISSIFAFSSLVFAMHLVFMSLGFSLVSGHAVSLP
                     YFGRESSISIFNSAATKLFFNSAIFWVLKLSAFFIGSIMVSALFLIGIAASALFC
                     AVLAGKKVVPEIFNPGAIFNIGLIGLWILKVSVKFSFLIGSLIVIIIGWL
                     IAFIKLAVDFFFKSMSIFFILIFIGAVALLSKAVKASEELLIIVGLIKKFF
                     KLIESELAVKLGRKLPEFSEALAKSLAKEVKTGLFS"
     misc_feature    21660..22000
                     /gene "H042H10.5"
                     /number 1
```



```
            are available and the accession number will be preserved
            * Note: This is a "working draft" sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="[illegible]"
                     /mol_type="Nippostrongylus"
                     /db_xref="taxon:[illegible]"
                     /clone_lib="[illegible]"
                     /note="[illegible]"
BASE COUNT    [illegible]
ORIGIN


Submitted (xx-[illegible]-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
Attributes were identified using ReportMaker:
Okubo, A.A.A. & Strong, M. (c)1996-1997
http://[illegible].wi.mit.edu:[illegible]/~[illegible]/ReportMaker.html
                  Genome Center
      Center: Whitehead Institute / MIT Center for Genome Research
      Center code: WIBR
      Web site: http://www.genome.wi.mit.edu
      Contact: sequence-submissions@genome.wi.mit.edu
                  Project Information
      Center project name: [illegible]
      Center clone name: [illegible]
                  ...
* NOTE: This record contains [illegible] individual
* sequenced reads that have not been assembled into
* contigs. Runs of Ns are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low pass sequence sampling is useful for
* identifying clones that may be gene rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*       1     721: contig of 721 bp in length
*     722  [illegible]: gap of     [illegible] bp
*     [illegible]  [illegible]: contig of 718 bp in length
*     [illegible]  [illegible]: gap of     200 bp
*     [illegible]  2411: contig of 685 bp in length
*     2412  2611: gap of     200 bp
*     2612  3101: contig of 690 bp in length
*     3102  3201: gap of     100 bp
*     3202  3916: contig of 715 bp in length
*     3917  4016: gap of     100 bp
*     4017  4703: contig of 687 bp in length
*     4704  4803: gap of     100 bp
*     4804  5518: contig of 715 bp in length
*     5519  5618: gap of     100 bp
*     5619  6332: contig of 714 bp in length
*     6333  6432: gap of     100 bp
*     6433  7137: contig of 705 bp in length
*     7138  7237: gap of     100 bp
*     7238  7941: contig of 704 bp in length
*     7942  8041: gap of     100 bp
*     8042  8744: contig of 703 bp in length
*     8745  8844: gap of     100 bp
*     8845  9547: contig of 703 bp in length
*     9548  9647: gap of     100 bp
*     9648  [illegible]: contig of 706 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 702 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 707 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 706 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 720 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 709 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 708 bp in length
*     [illegible]  [illegible]: gap of     100 bp
*     [illegible]  [illegible]: contig of 675 bp in length
*     [illegible]  [illegible]: gap of     100 bp
```





```
            [illegible] Wang,Y. [illegible] and [illegible]
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]) Chinese National Human Genome Center at
            Shanghai, [illegible] Rd, [illegible] Hi-Tech Park, Pudong,
            Shanghai [illegible] P. R. China
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="hypothalamus"
     CDS             [illegible]
                     /note="ARPP-21"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="cyclic AMP-regulated phosphoprotein"
                     /protein_id="[illegible]"
                     /db_xref="[illegible]"
                     /translation="MSE[illegible]"
BASE COUNT     [illegible]
ORIGIN

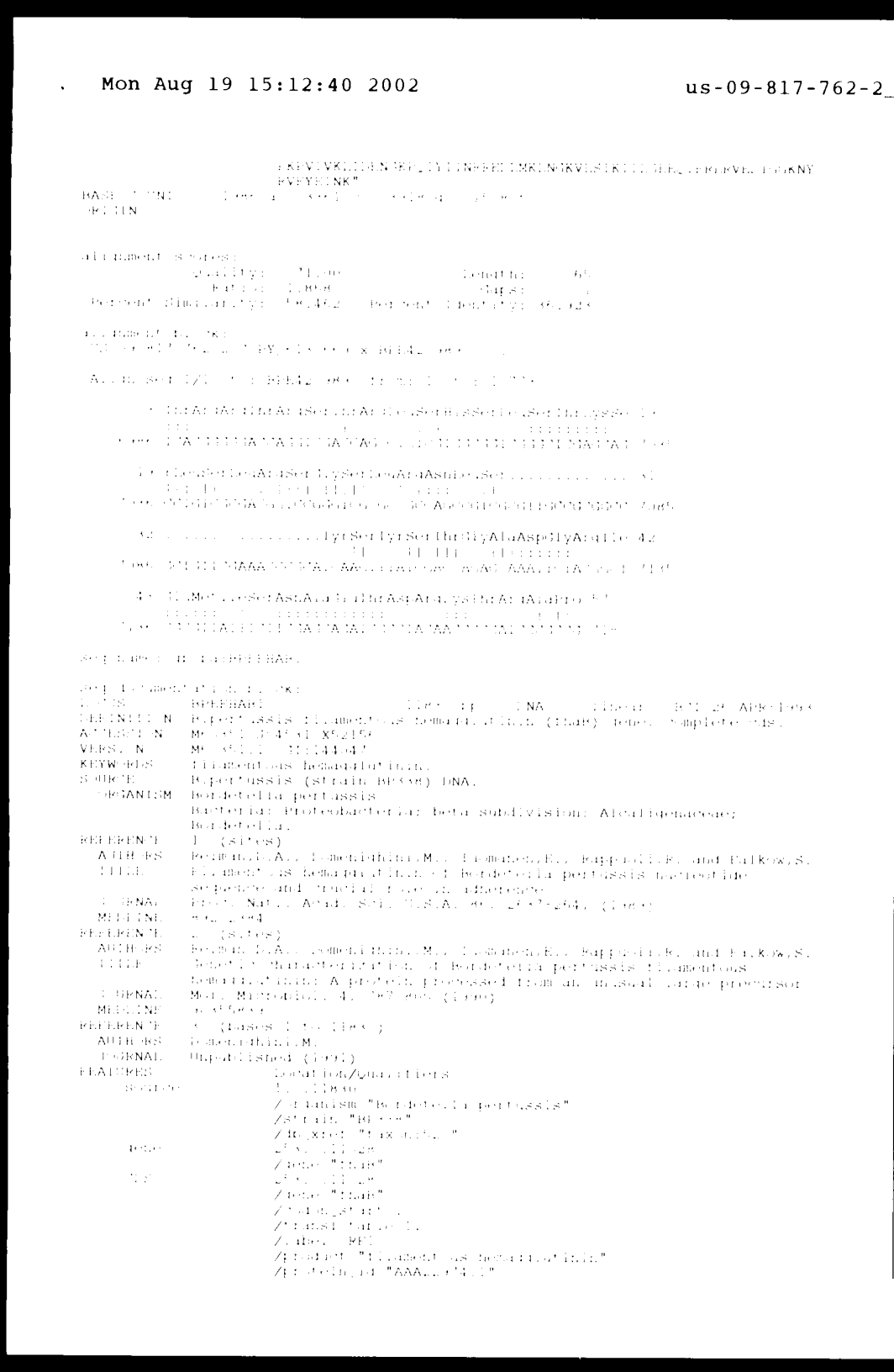
Alignment Scores:
Quality:          [illegible]        Length:        [illegible]
Ratio:            [illegible]        Gaps:          [illegible]
Percent Similarity: 66.667   Percent Identity: [illegible]

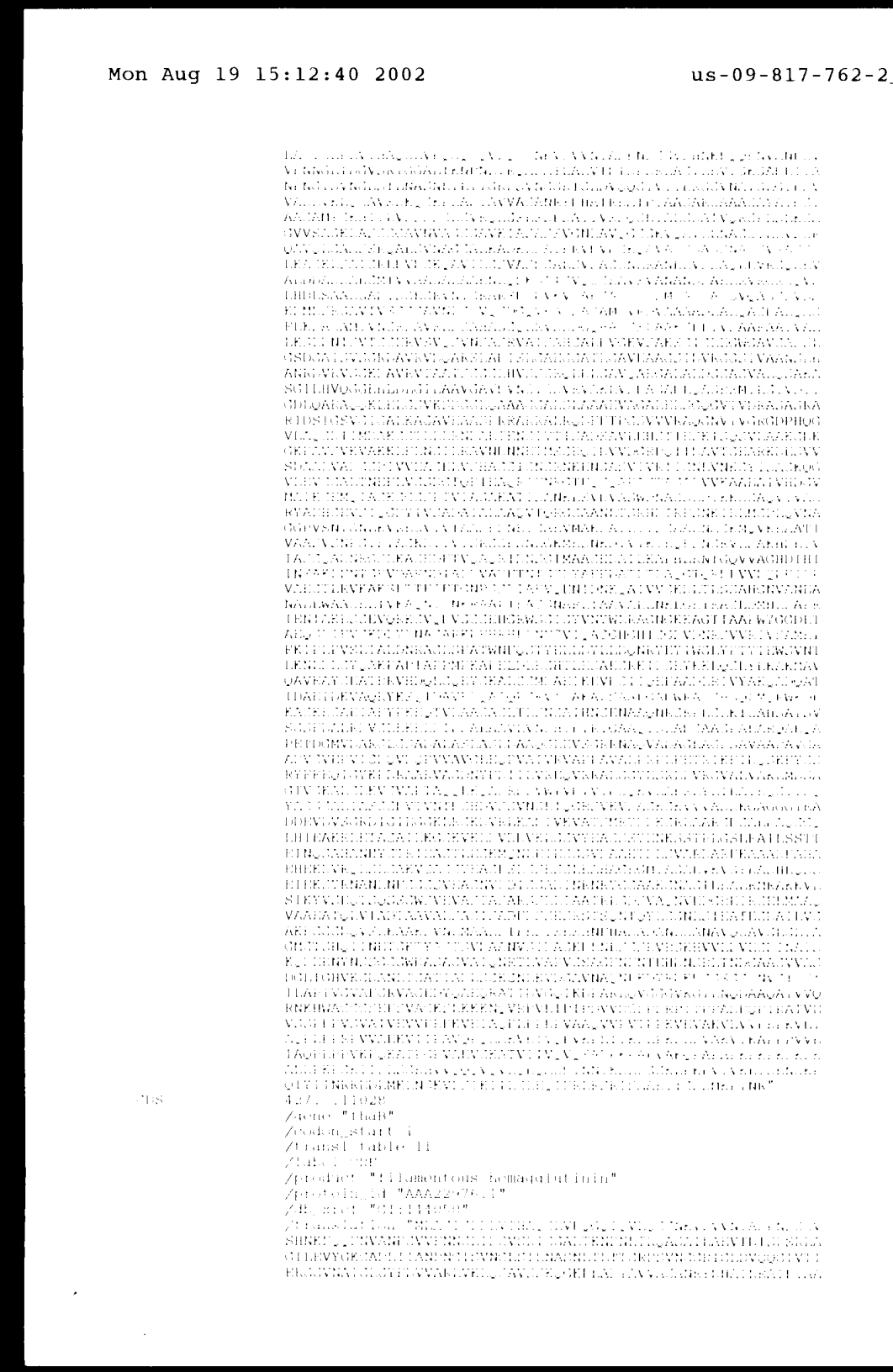
Alignment Block:
US-09-817-762-2-COPY-613-669 x AF[illegible]

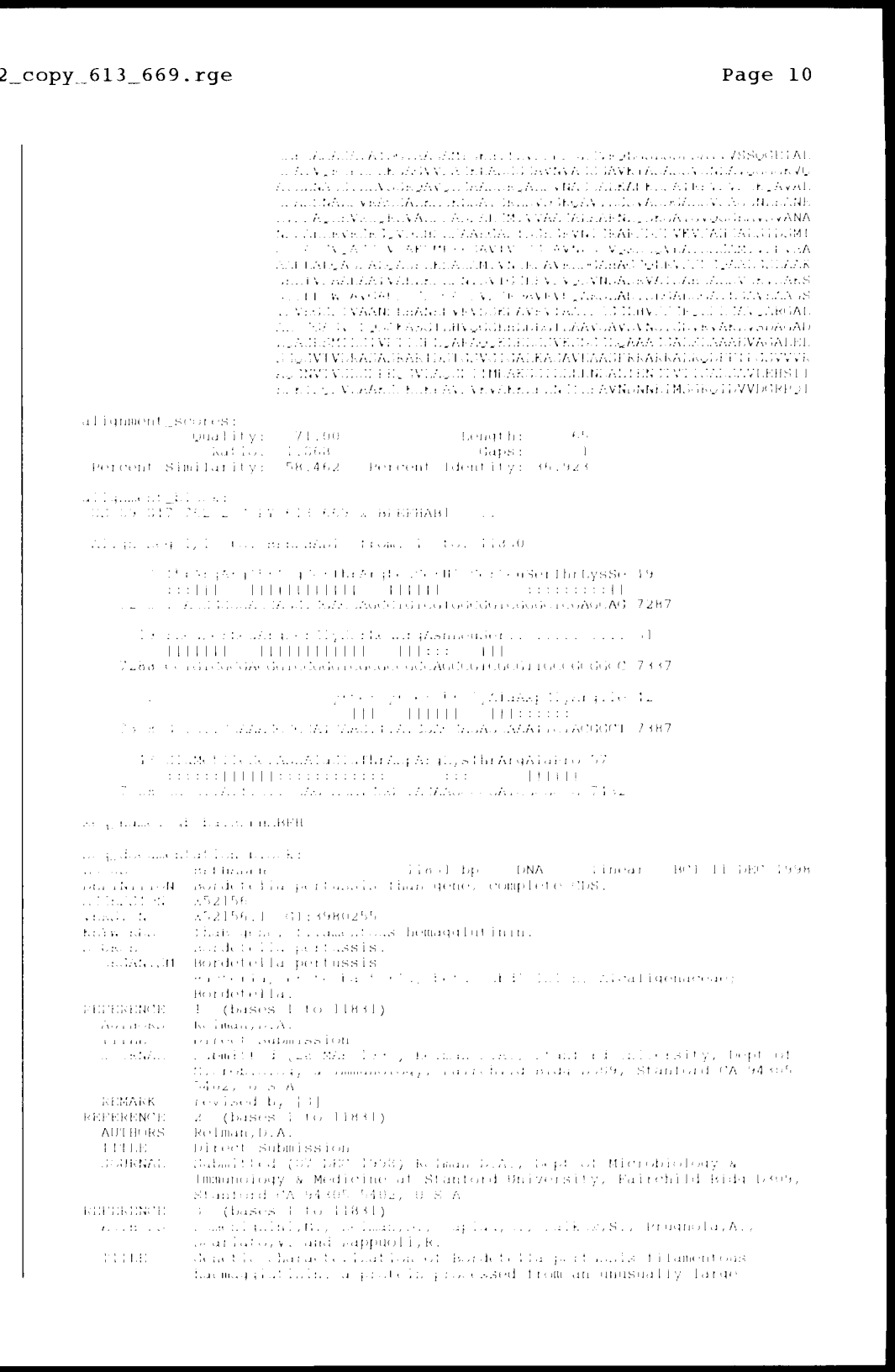
[illegible]

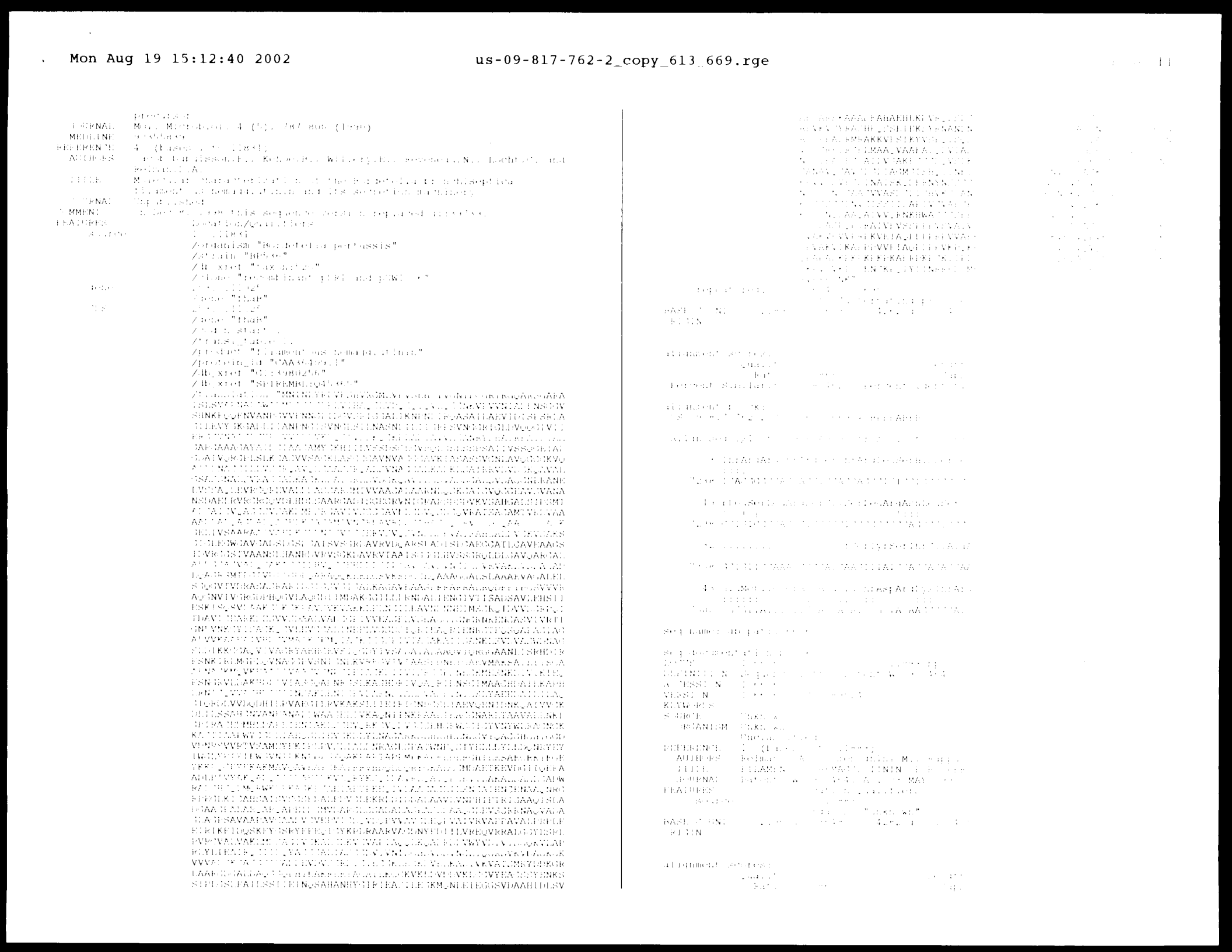
Sequence name: [illegible]

Sequence Information Block:
LOCUS       [illegible]      [illegible] bp    DNA     linear   BCT 28-NOV-[illegible]
DEFINITION  Bordetella pertussis fhaB2 gene for FHA protein.
ACCESSION   AJ[illegible]
VERSION     AJ[illegible]  GI:[illegible]
KEYWORDS    FHA protein; fhaB2 gene.
SOURCE      Bordetella pertussis.
  ORGANISM  Bordetella pertussis
            Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
            Bordetella.
REFERENCE   1  (sites)
  AUTHORS   van Loo,I.H., Heuvelman,K.J., King,A.J. and Mooi,F.R.
  TITLE     Multilocus sequence typing of Bordetella pertussis based on surface
            proteins: identification of epidemic types
  JOURNAL   Unpublished
REFERENCE   2  (bases [illegible])
  AUTHORS   van Loo,I.H.
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]-NOV-[illegible]) van Loo I.H., Lab. for Infectious Diseases,
            Research [illegible] Nat. Inst. for Public Health and Environ., postbox [illegible],
            BA Bilthoven, NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Bordetella pertussis"
                     /db_xref="taxon:520"
     gene            1..10773
                     /gene="fhaB2"
     CDS             1..10773
                     /gene="fhaB2"
                     /codon_start=1
                     /transl_table=11
                     /product="FHA protein"
                     /protein_id="CAD12824.1"
                     /db_xref="GI:17154561"
                     /translation="MNTNLYRLVFSHVRGMLVPVSEHCTVGNTFCGRRRLAAGLA[illegible]
                     [illegible]
```









```
[illegible]

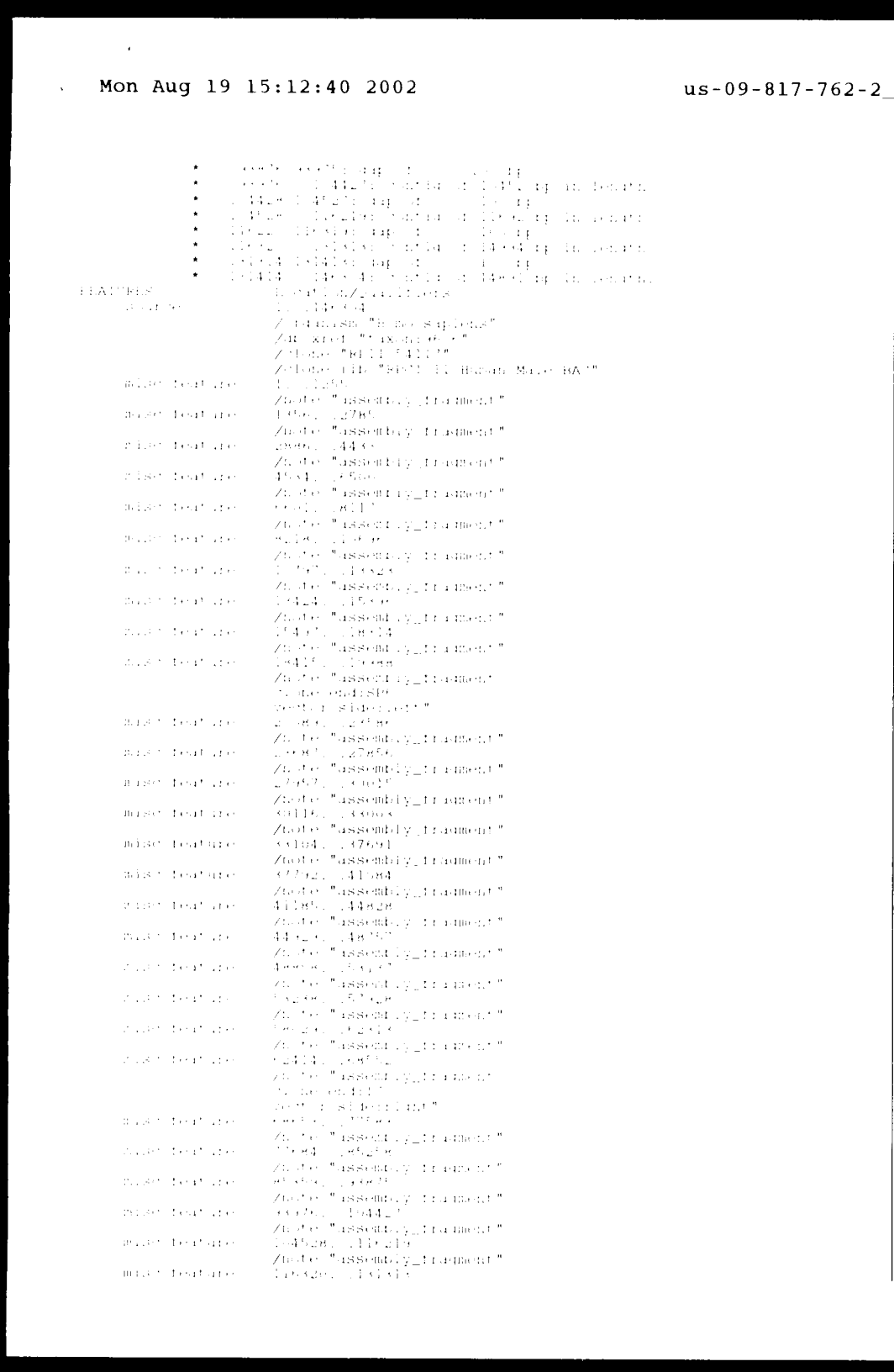
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
  ORGANISM
REFERENCE
  AUTHORS
  TITLE
  JOURNAL
REFERENCE
  AUTHORS
  TITLE
  JOURNAL
COMMENT
  [illegible]

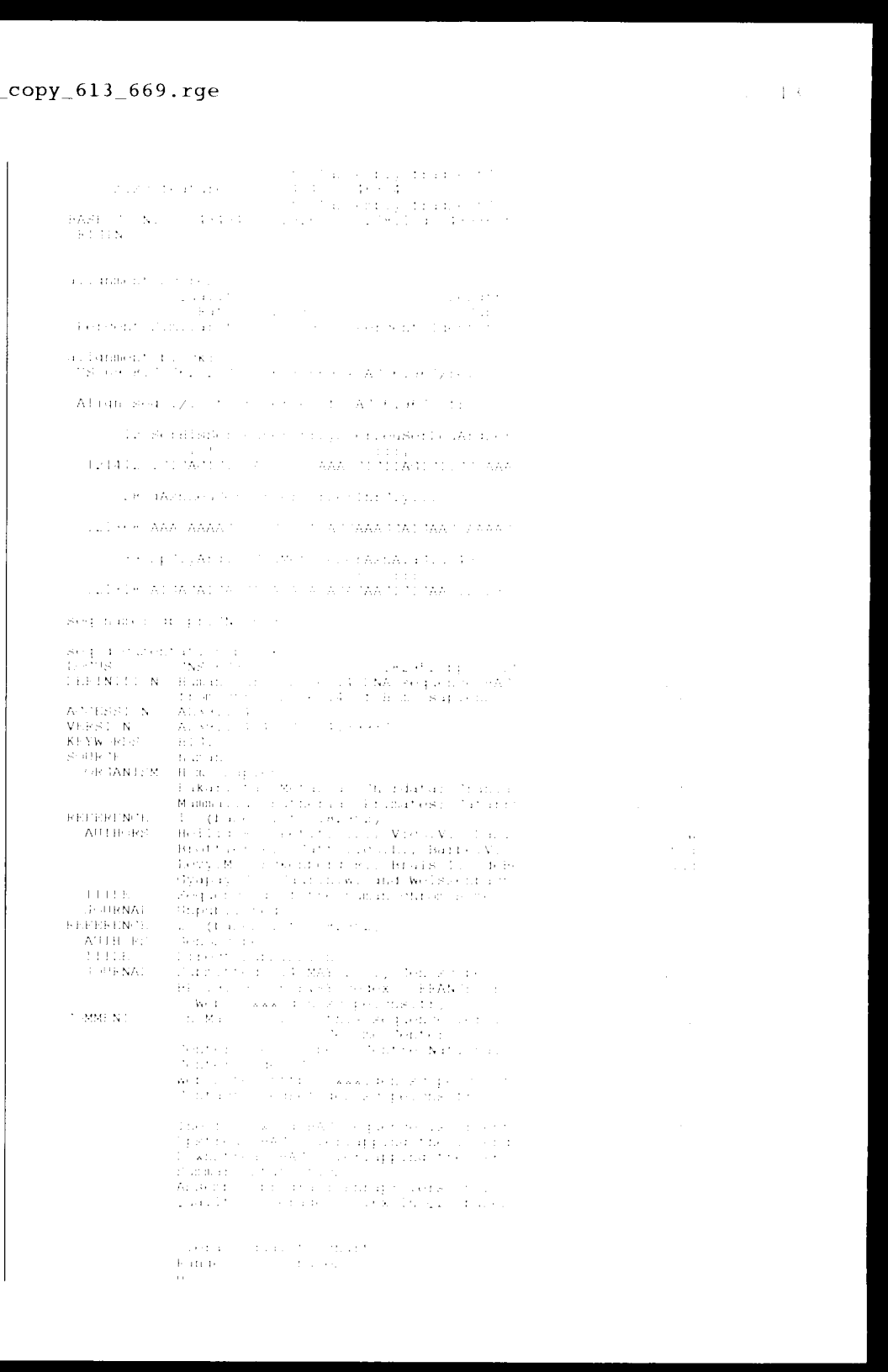
seq name: qb|btq|A*06|067

seq documentation block:
LOCUS       A*06:067                146304 bp    DNA     linear   HTG 25-MAY-2000
DEFINITION  Homo sapiens clone RPCI-94117, WORKING DRAFT SEQUENCE, 29 unordered
            pieces.
ACCESSION   A*06:067
VERSION     A*06:067.x  GI:[illegible]
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146304)
  AUTHORS   Barrett,B., Linton,L., Nusbaum,C., and Lander,E.
  TITLE     Homo sapiens, c, clone RPCI-94117
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146304)
  AUTHORS   Barrett,B., Linton,L., Nusbaum,C., Lander,E., Abraham,K., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastian,V., Beck,S.,
            Bentolila,S., Bookbinder,E., Brown,A., Burkett,T.,
            Campagnoni,A., Castle,A., Change,Y., Colamarino,M., Collins,S.,
            Cuppernull,A., Decker,E., DeArellano,K., Dewar,K., Eliaz,Y.,
            Foster,S., Gonzalez,M., Hayles,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Gallaher,J., Gardyna,S., Gitade,S., Goyette,M., Graham,I.,
            Grand Pierre,N., Grant,G., Hanos,B., Heaford,A., Horton,L.,
            Howland,J.C., Hsiao,L., Johnson,E., Jones,C., Kamal,I., Karatas,A.,
            Kliot,J., LaRosque,K., Lamarosa,K., Landers,F., Lebovsky,J.,
            Levine,K., Liem,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McCann,P., McClure,A., McKernan,K., Mikeladze,I.,
            McElrine,J., Monsen,L., Millers,P., Miranda,C., Mistry,V., Morrow,J.,
            Murphy,L., Naylor,J., Norman,C.E., O'Connor,L., O'Donnell,P.,
            O'Neil,L., Olivar,L.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Roane,P., Rothman,L.,
            Ko,A., Santos,K., Schauer,S., Severy,P., Spencer,B.,
            Stanke Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Jostrzeb,S., Trevarrow,L., Ulrel,A., Waters,M., Iriqilio,J.,
            Vassiliev,H., Vliet,R., V...A., Wilson,R., Wu,X., Wylie,[illegible]W...,
            Yoder,L., Zaibout,J., Zimmer,A. and Eudy,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2[illegible] 2001 this sequence version replaced gi:[illegible].
            Also, reports with Unpublished Gaps Report Markers:
            [illegible] Acids Res. 9 Suppl. 1:[illegible] (1981)
            [illegible]http://[illegible].bcm.tmc.edu/HGP/ReportMasker.htm.
                        Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www.genome.wi.mit.edu

            Contact: Sequence Submissions genome-www@mit.edu
                     Project Information
            Center project name: [illegible]
            Center clone name: 941 1 7
                     Summary Statistics
            Sequencing vectors: MIS: M77811; [illegible] reads
            Cloning vectors: bac-terminator BD type: [illegible] reads
            Assembly program: Phrap, version [illegible]
            Consensus quality: 18217 bases at least q40
            Consensus quality: 19941 bases at least q30
            Consensus quality: 14267 bases at least q20
            Insert size: 145[illegible] coverage 5p
            Insert size: 145[illegible] 5.8 of contigs
            Gaps: [illegible], coverage: 5.8 in q20 bases, coverage 5p
            Gaps: [illegible], coverage: 5.9 in q20 bases, 5.8 of contigs

* NOTE: This is a "working draft" sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1      1255: contig of 1255 bp in length
*    1256   1355: gap of      100 bp
*    1356      2785: contig of 1430 bp in length
*    2786   2885: gap of      100 bp
*    2886      4133: contig of 1248 bp in length
*    4134   4533: gap of      400 bp
*    4534      6500: contig of 1967 bp in length
*    6501   6600: gap of      100 bp
*    6601      8117: contig of 1517 bp in length
*    8118   8217: gap of      100 bp
*    8218      10696: contig of 2479 bp in length
*   10697  10796: gap of      100 bp
*   10797      13323: contig of 2527 bp in length
*   13324  13423: gap of      100 bp
*   13424      15396: contig of 1973 bp in length
*   15397  15496: gap of      100 bp
*   15497      18310: contig of 2814 bp in length
*   18311  18414: gap of      104 bp
*   18415      19988: contig of 1574 bp in length
*   19989  20088: gap of      100 bp
*   20089      23586: contig of 3498 bp in length
*   23587  23686: gap of      100 bp
*   23687      27856: contig of 4170 bp in length
*   27857  27956: gap of      100 bp
*   27957      30015: contig of 2059 bp in length
*   30016  30115: gap of      100 bp
*   30116      33003: contig of 2888 bp in length
*   33004  33103: gap of      100 bp
*   33104      37[illegible]: contig of 4569 bp in length
*   37[illegible]  37791: gap of      100 bp
*   37792      41084: contig of 3293 bp in length
*   41085  41184: gap of      100 bp
*   41185      44828: contig of 3644 bp in length
*   44829  44928: gap of      100 bp
*   44929      48757: contig of 3829 bp in length
*   48758  48857: gap of      100 bp
*   48858      53185: contig of 4328 bp in length
*   53186  53285: gap of      100 bp
*   53286      57928: contig of 4643 bp in length
*   57929  58028: gap of      100 bp
*   58029      62[illegible]: contig of 4293 bp in length
*   [illegible]  [illegible]: gap of      100 bp
*   [illegible]      [illegible]: contig of 6136 bp in length
*   [illegible]  [illegible]: gap of      100 bp
*   [illegible]      [illegible]: contig of 8041 bp in length
*   [illegible]  [illegible]: gap of      100 bp
*   [illegible]      [illegible]: contig of 7575 bp in length
*   [illegible]  [illegible]: gap of      100 bp
*   [illegible]      [illegible]: contig of 8517 bp in length
```





```
          [illegible]
          [illegible]
          [illegible]
          30 - 39 :     164
          40 - 49 :    1446
          [illegible]    6582
          [illegible]    7218
          [illegible]
          [illegible]
          [illegible]

          Percentage of bases with a quality value > 4[illegible]
FEATURES             Location/Qualifiers
     source          1..182962
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "14"
                     /clone "R-413E8"
                     /clone_lib "RPCI-11"
     STS             [illegible]
                     /note "matching EMBL:[illegible]
                     RHdb:RH[illegible]14
                     RHdb:RH[illegible]44
                     dbSTS:STS[illegible]
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT     [illegible] a  35426 c  33884 g  [illegible] t
ORIGIN


Alignment Scores:
            Quality:    [illegible]              Length:      41
            Ratio:   2.[illegible]               Gaps:         2
Percent Similarity: [illegible]   Percent Identity: 4[illegible]

Alignment to Query:
  Seq 817 762 2 COPY 613 669 x CNS06780/rev [illegible]

Alignment seq 1/1 to reverse of: CNS06780  from: 1  to: 182962

     12 SerHisSerLeuSerThrLysSerLeuSerLeuArgSerGlySerLeuAr 28
                      [illegible]
  [illegible] [illegible]AAAT[illegible]AAA[illegible]AAA[illegible]

     28 gAsnLeuSerTyrSerTyrSerThrGly.....................AlaA 39
          [illegible]
  [illegible] AA[illegible]

     39 spGlyArgIleGluMetIleSerAsnAlaGlu 49
        ||         |           ,|:::||'::|||
  [illegible]

Seq_name: gb|[illegible]|CNS0[illegible]V

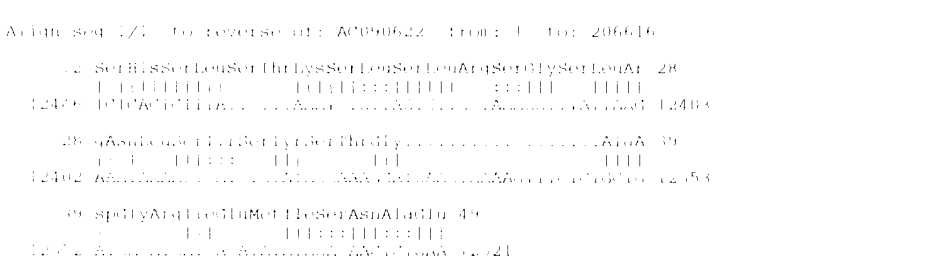
Seq Documentation Block:
LOCUS       CNS0[illegible]V          206413 bp    DNA     linear   PRI 04-JUL-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-355I22 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL390816
VERSION     AL390816.4  GI:14625742
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206413)
  AUTHORS   [illegible], Petit,J.L., Vico,V., Basil[illegible], Robert,[illegible], Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206413)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]-JUL-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jul 6, 2001 this sequence version replaced gi:[illegible]
                         Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr

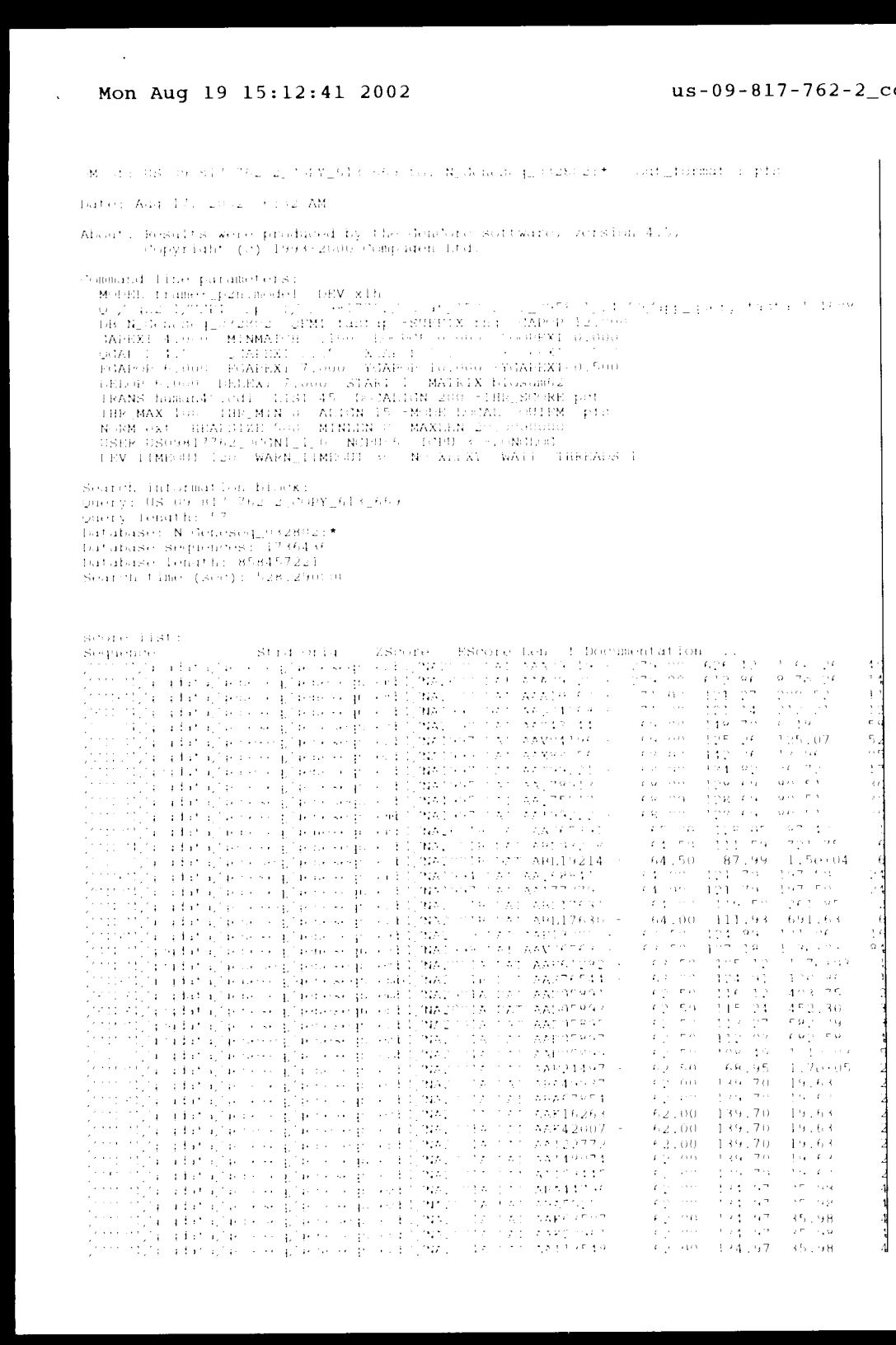
            The following BAC sequence is oriented from the 17[illegible] the SP[illegible] end
            Upstream BAC (overlapping the 17 end) : [illegible]227K2
            Downstream BAC (overlapping the SP6 end) : R-113E8 (AC AL[illegible])
                         Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.95x in Q20 bases; sum of contigs

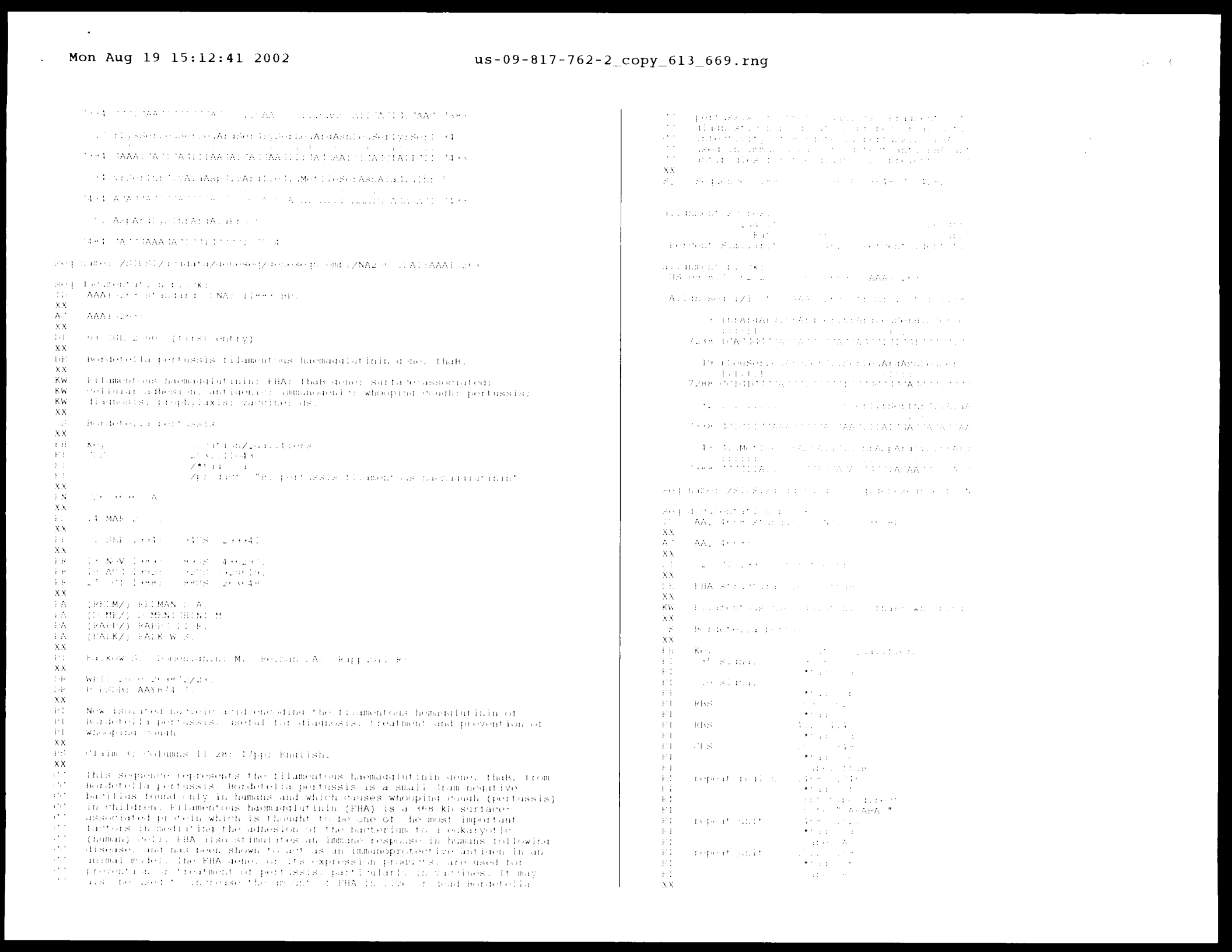
            Detail quality chart :
            Range   :   Bases
                    :
            [illegible] - 9 :
            10 - 19 :
            20 - 29 :         1
            30 - 39 :        39
            40 - 49 :      1216
            50 - 59 :      [illegible]
            60 - 69 :      [illegible]
            70 - 79 :     [illegible]
            80 - 89 :     [illegible]
            90 - [illegible] :    [illegible]

            Percentage of bases with a quality value > 40 : [illegible]
FEATURES             Location/Qualifiers
     source          1..206413
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "14"
                     /clone "R-355I22"
                     /clone_lib "RPCI-11"
     STS             [illegible]..[illegible]547
                     /note "matching EMBL:[illegible]
                     RHdb:RH53549
                     dbSTS:STS16033
                     Identified using the e-PCR software (G. Schuler)"
     STS             49495..49795
                     /note "matching EMBL:G03447
                     RHdb:RH34715
                     dbSTS:STS8202
                     Identified using the e-PCR software (G. Schuler)"
     STS             74[illegible]..74681
                     /note "matching EMBL:Z44[illegible]24
                     RHdb:RH[illegible]
                     dbSTS:STS[illegible]
                     Identified using the e-PCR software (G. Schuler)"
     STS             76429..76560
                     /note "matching EMBL:F02008
                     RHdb:RH10042
                     dbSTS:STS20999
                     Identified using the e-PCR software (G. Schuler)"
     STS             9[illegible]..9[illegible]
                     /note "matching EMBL:AA[illegible]
                     RHdb:RH[illegible]
                     dbSTS:STS[illegible]
                     Identified using the e-PCR software (G. Schuler)"
     STS             94[illegible]..94661
                     /note "matching EMBL:[illegible]
                     RHdb:RH47362
                     dbSTS:STS40422
                     Identified using the e-PCR software (G. Schuler)"
     STS             1[illegible]..198118
                     /note "matching EMBL:[illegible]
```

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 296616)
AUTHORS
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 296616)
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN







```
RN   [1]
XX
RA   Ramsburg H.H.;
XX
RT   "DNA MAY [illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [2]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [3]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [4]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [5]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [6]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [7]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [8]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [9]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [10]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [11]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [12]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [13]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [14]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [15]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [16]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [17]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [18]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [19]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [20]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [21]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [22]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [23]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [24]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [25]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [26]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [27]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [28]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [29]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [30]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [31]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [32]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [33]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [34]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [35]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [36]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [37]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [38]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [39]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [40]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [41]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [42]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [43]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [44]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [45]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [46]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [47]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [48]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [49]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [50]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [51]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [52]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [53]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [54]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [55]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [56]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [57]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [58]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [59]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [60]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [61]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [62]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [63]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [64]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [65]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [66]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [67]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [68]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [69]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [70]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [71]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [72]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [73]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [74]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [75]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [76]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [77]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [78]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [79]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [80]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [81]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [82]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [83]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [84]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [85]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [86]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [87]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [88]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [89]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [90]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [91]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [92]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [93]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [94]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [95]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [96]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [97]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [98]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [99]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [100]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [101]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [102]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [103]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [104]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [105]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [106]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [107]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [108]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [109]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [110]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [111]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [112]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [113]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [114]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [115]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [116]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [117]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [118]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [119]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [120]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [121]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [122]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [123]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [124]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [125]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [126]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [127]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [128]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [129]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [130]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [131]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [132]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [133]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [134]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [135]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [136]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [137]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [138]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [139]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/GenBank/DDBJ databases.
RL   [illegible] [illegible], [illegible], [illegible], [illegible], [illegible].
XX
RN   [140]
XX
RA   [illegible];
XX
RT   "[illegible]";
XX
RL   Submitted (06-MAY-[illegible]) to the EMBL/
```

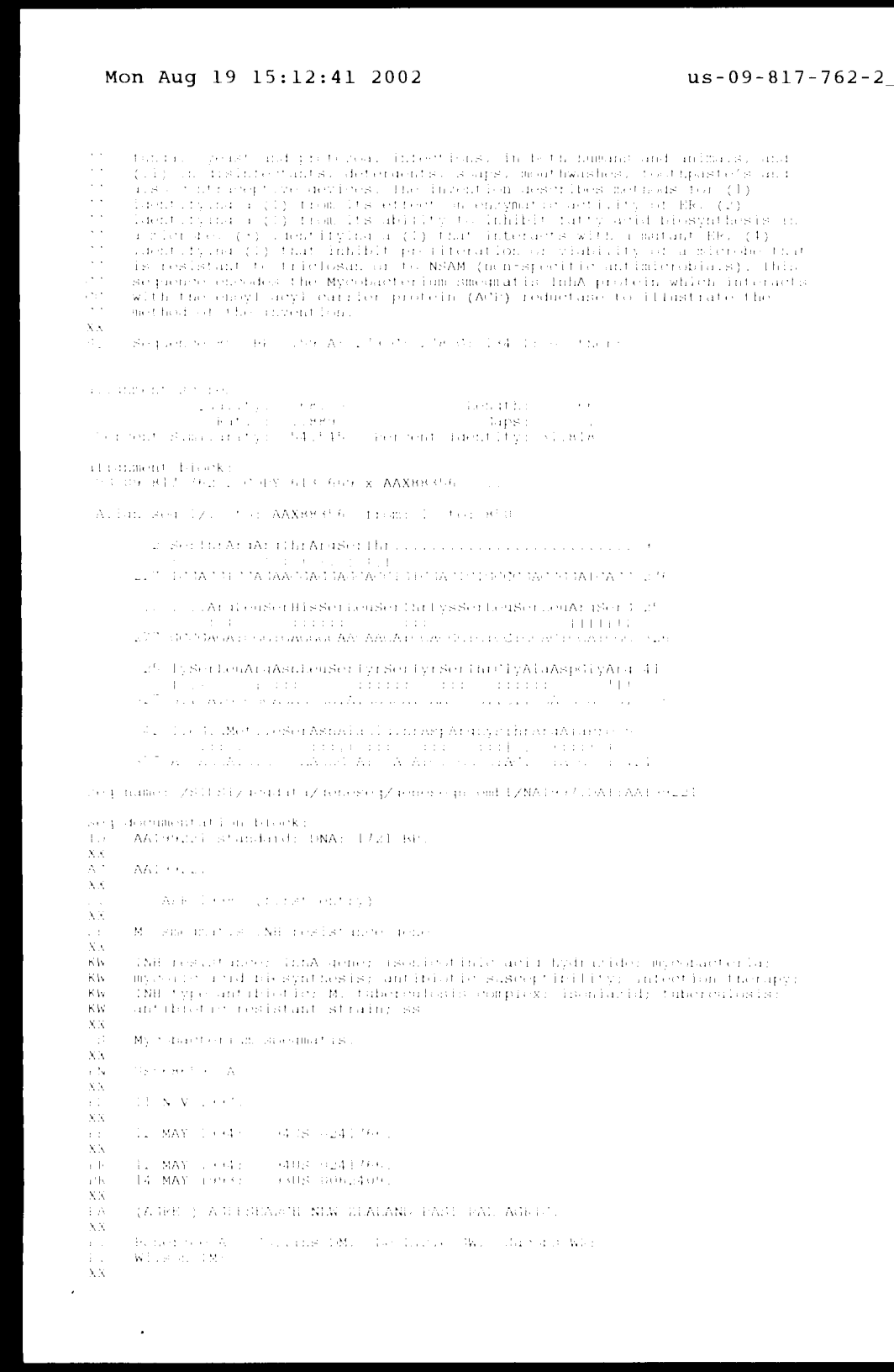
```
Seq name: /SIDS/.../AAV04196
Seq documentation block:
ID   AAV04196 standard; DNA; 5252 BP.
XX
AC   AAV04196;
XX
DT   14 APR 1998  (first entry)
XX
DE   Trans feruloyl SCoA hydratase/vanillin:NAD+ oxidoreductase DNA
XX
KW   trans feruloyl SCoA hydratase; vanillin production; vanillin pod;
KW   vanillic acid production; food flavouring;
KW   vanillin:NAD+ oxidoreductase; vanillin dehydrogenase; ss.
XX
OS   Pseudomonas fluorescens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2472..3302
FT                   /*tag= a
FT                   /product= "trans feruloyl SCoA_hydratase"
FT   CDS             3804..5252
FT                   /*tag= b
FT                   /product= "vanillin:NAD+_oxidoreductase"
XX
PN   WO9735999-A1.
XX
PD   02-OCT-1997.
XX
PF   24-MAR-1997;   97WO-GB00809.
XX
PR   25-MAR-1996;   96GB-0006187.
XX
PA   (BRIT ) BRITISH RES INST.
XX
PI   Gasson MJ,  Narbad A,  Rhodes MJC,  Walton NJ;
XX
DR   WPI; 1997-469549/43.
DR   P-PSDB; AAW41560, AAW41561.
XX
PT   Production of vanillin from trans ferulic acid using 3 enzymes
PT   optionally followed by oxidation to vanillic acid, useful as a food
PT   flavouring
XX
PS   Example 1; Fig 12; 127pp; English.
XX
CC   The present sequence encodes Pseudomonas fluorescens trans feruloyl
CC   SCoA hydratase/vanillin:NAD+ oxidoreductase (vanillin dehydrogenase),
CC   which was used in the development of a novel method for the
CC   production of vanillin.
CC   The method comprises exposing trans ferulic acid, or its salts, to
CC   trans ferulate:CoASH ligase, trans feruloyl SCoA hydratase and
CC   4-hydroxy-3-methoxyphenyl beta hydroxypropionyl-S-CoA cleavage
CC   activities. A similar novel method in which vanillin is converted
CC   to vanillic acid, or its salts, utilises a 4th enzyme, especially
CC   vanillin dehydrogenase. Vanillin is a well known food flavouring,
CC   and vanillic acid can be used to make oligomers, or as a monomer
CC   for production of polyesters. By expressing vanillin in plants or
CC   food grade microorganisms, vanillin is generated in situ during
CC   preparation of fermented foods and beverages. Transgenic plants,
CC   particularly food plants, may be able to produce other
CC   flavour/aroma compounds in addition to vanillin. This method is
CC   less expensive than extraction from Vanilla pods, and generates a
CC   natural product as only enzymes are used. The product is very
CC   similar to natural Vanilla pod extract.
XX
SQ   Sequence 5252 BP; ... A; ... C; ... G; ... T; ... other;

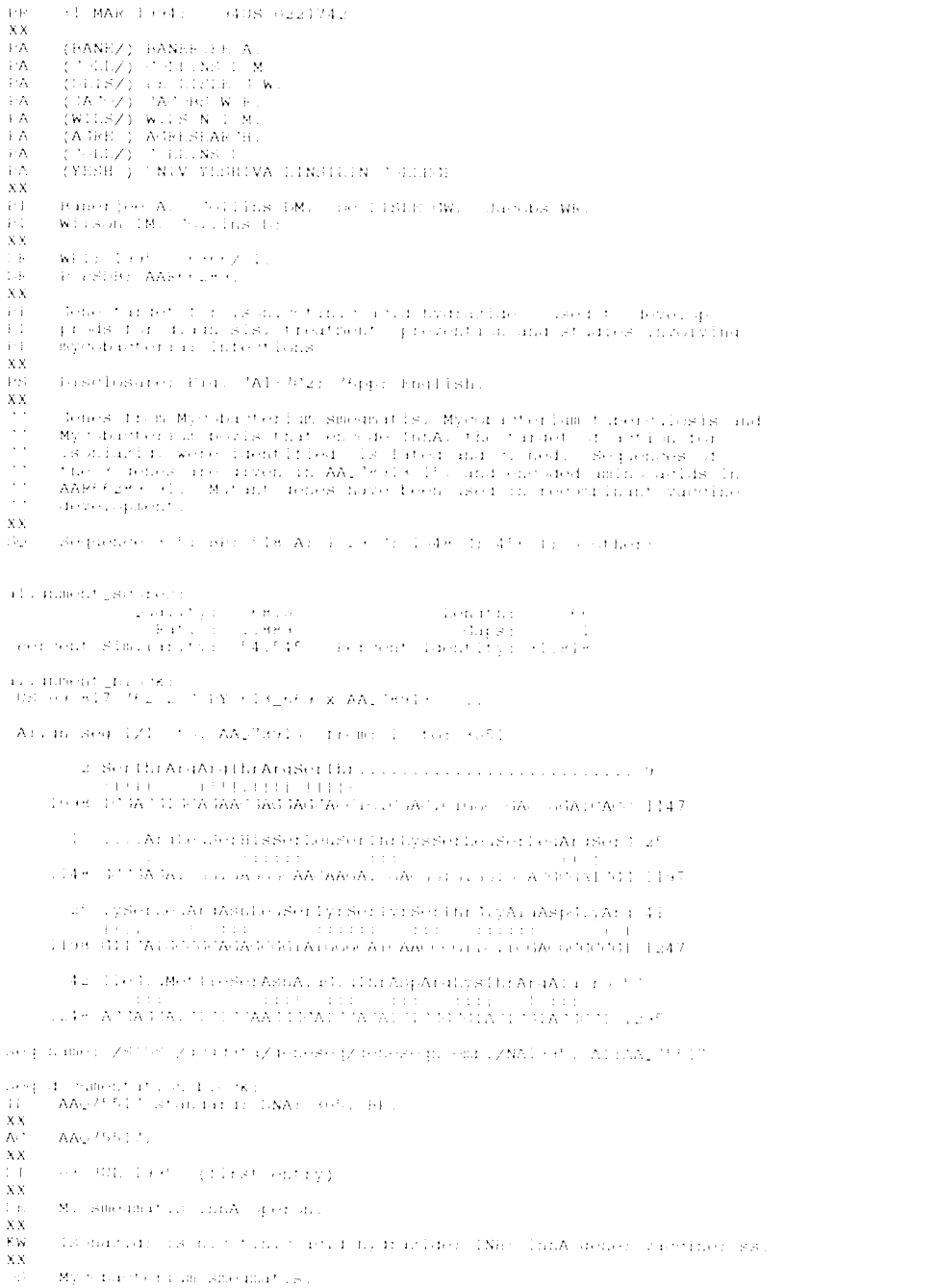
Alignment Scores:
      Quality:   ...                Length:     4..

Percent Similarity: ...
Alignment Length: ...
US-09-817-762-2 (1-...) x AAV04196 ...

Align seq 1/...

AAX86456 ...
Seq documentation block:
ID   AAX86456 standard; DNA; ... BP.
XX
AC   AAX86456;
XX
DT   ... SEP 1999 (first entry)
XX
DE   M... ...
XX
KW   ...
KW   ...
KW   ...
KW   ...
KW   ...
XX
OS   My...
XX
FH   Key             Location/Qualifiers
FT   CDS             ...
FT                   /*tag= a
FT                   /...="...A"
XX
PN   WO99...-A1.
XX
PD   ...
XX
PF   22-JAN-1999; ...
XX
PR   2?-JAN-1998; ...
PR   2?-JAN-1998; ...
XX
PA   (UYJO ) ...
XX
PI   ...
XX
DR   WPI; 1999-4...
DR   P-PSDB; AAY...
XX
PT   Identifying ...
PT   ...
PT   and detergent...
XX
PS   ...
XX
CC   ...
```





```
XX
KW   INH resistance; InhA gene; isonicotinic acid hydrazide; mycobacteria;
KW   mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;
KW   INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KW   [illegible] resistant strain; ds.
XX
OS   Mycobacterium smegmatis.
XX
PN   [illegible] A.
XX
PD   [illegible] NOV [illegible].
XX
PF   [illegible] MAY [illegible]
XX
PR   [illegible] MAY [illegible]
PR   [illegible] MAY [illegible]
XX
PA   (AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRI[illegible]
XX
PI   Banerjee A, Collins DM, De Lisle GW, Jacobs WR,
PI   Wilson TM;
XX
DR   WPI; [illegible]
XX
PT   Nucleic acids encoding mycobacterial InhA proteins - associated with
PT   isoniazid susceptibility and resistance
XX
PS   Disclosure; [illegible] 38; 55pp; English.
XX
CC   This sequence represents the coding sequence for a mycobacterial InhA
CC   protein. The encoded protein is involved in mycolic acid biosynthesis.
CC   The nucleic acids are useful in assessing the susceptibility of various
CC   strains of the M. tuberculosis complex to isoniazid (INH) type
CC   antibiotics, as decoys and as antisense oligonucleotides to prevent the
CC   expression of polypeptides associated with isoniazid (also referred to as
CC   isonicotinic acid hydrazide) resistance, and for the expression of the
CC   proteins, which may also have use in immunoassays for the detection of
CC   INH resistant strains, in the determination of whether an INH type
CC   antibiotic may be effective against tuberculosis, and in the treatment of
CC   individuals for infection with these strains.
XX
SQ   Sequence [illegible] BP; [illegible] A; [illegible] C; [illegible] G; [illegible] T; [illegible] other;


Alignment Scores:
                    Quality: [illegible]          Length: [illegible]
                    Ratio: [illegible]            Gaps: [illegible]
  Percent Similarity: [illegible]    Percent Identity: [illegible]

Alignment Block:
 US-09-817-762-2_COPY_613_669 x AAP[illegible]

  Align seq 1/1 to: AAP[illegible] from: 1 to: [illegible]

       1 SerThrArgArgThrArgSerThr.......................... [illegible]

   [illegible] 1117

       [illegible] ...ArgLeuSerHisSerLeuSerThrLysSerLeuSerLeuArgSer [illegible]

   [illegible] 1197

       [illegible] LysSerLeuArgAsnLeuSerTyrSerTyrSerThrTyrAlaAspGlyArg [illegible]

   [illegible] 1247

       [illegible] MetLeuSerAsnAla[illegible]ThrAspArg[illegible]ThrArgAla[illegible]

   [illegible] 1295

Sequence /[illegible]/geneseq/geneseqp-emb1/NA20[illegible]:AAS[illegible]

Seq [illegible]
```

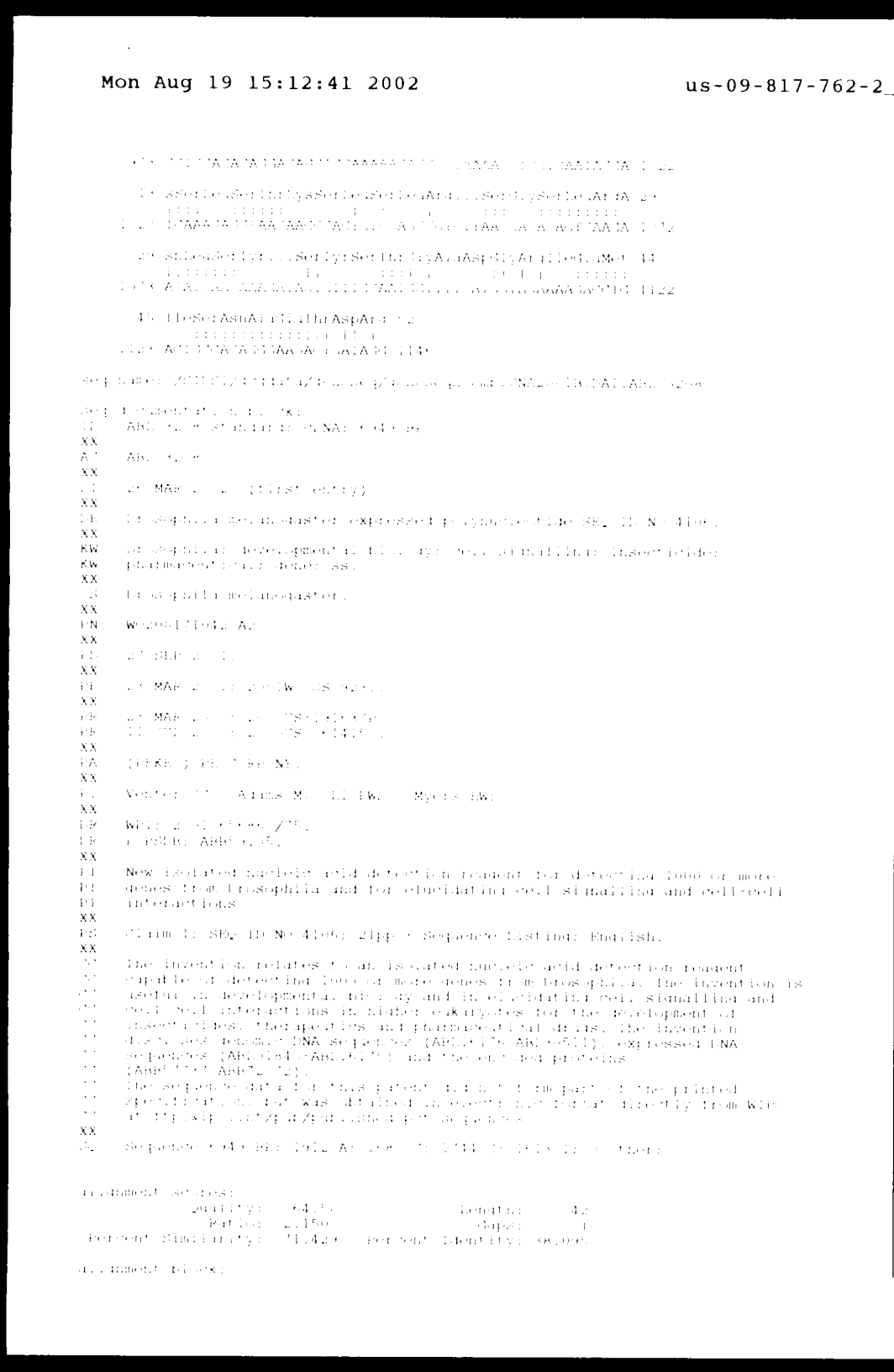
```
ID   AAS[illegible] standard; cDNA; 1665 BP.
XX
AC   AAS[illegible];
XX
DT   [illegible] (first entry)
XX
DE   DNA encoding novel human diagnostic protein #[illegible].
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO[illegible] A1.
XX
PD   [illegible]
XX
PF   30 MAR [illegible]; [illegible]W[illegible] US[illegible].
XX
PR   31 MAR 2000; 2000US-0540217.
PR   23 AUG 2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   [illegible] Y[illegible];
XX
DR   WPI; [illegible]/[illegible].
DR   P-PSDB; ABG[illegible]1204.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim [illegible]; SEQ ID NO [illegible]; [illegible]pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS84197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1665 BP; 380 A; 487 C; 521 G; 277 T; 0 other;


Alignment Scores:
                    Quality: [illegible]          Length: [illegible]
                    Ratio: [illegible]            Gaps: [illegible]
  Percent Similarity: [illegible]    Percent Identity: [illegible]

Alignment Block:
 US-09-817-762-2_COPY_613_669 x AAS[illegible]

  Align seq 1/1 to: AAS[illegible] from: 1 to: 1665

       1 [illegible]SerThrArgArgThrArgSerThrArgLeuSer.......... [illegible]
```

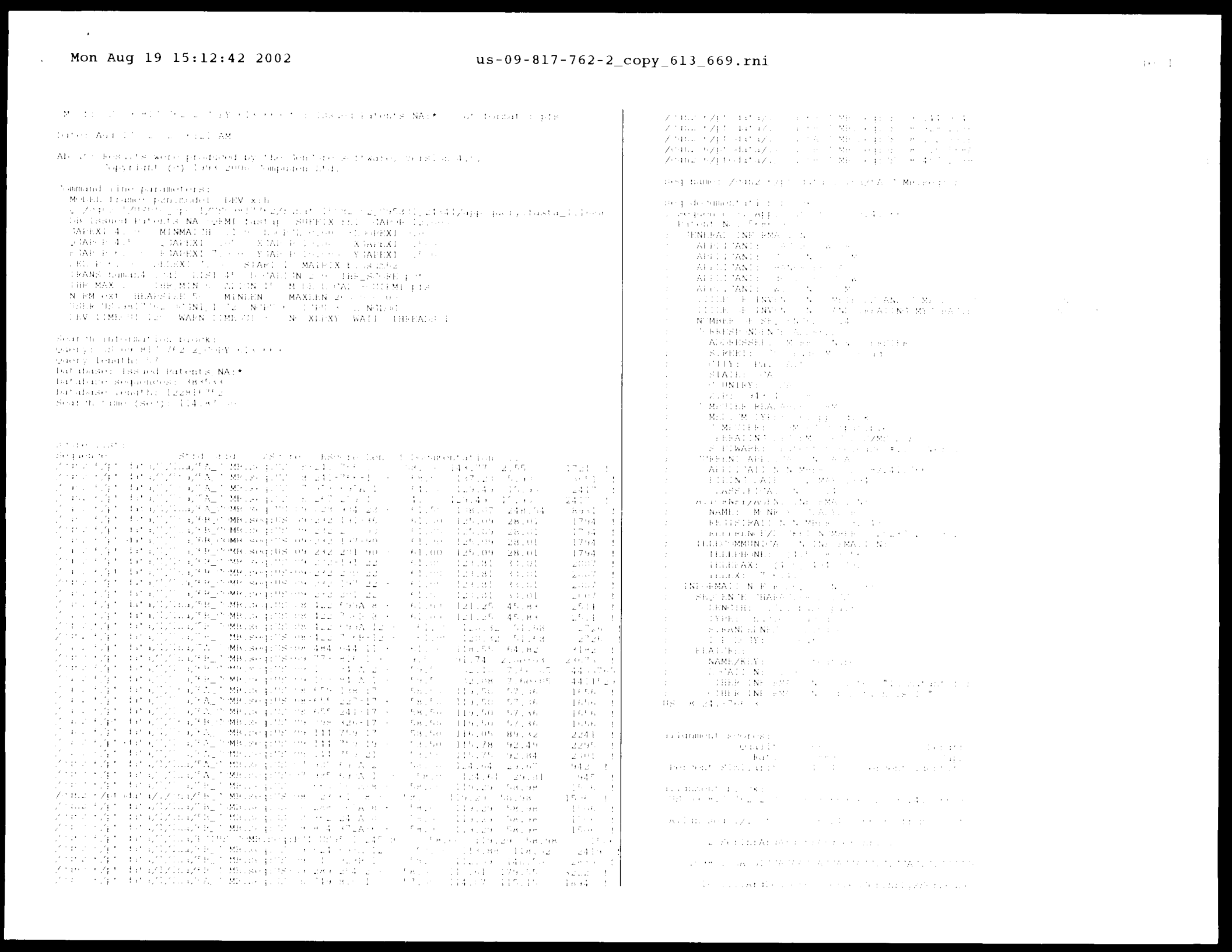


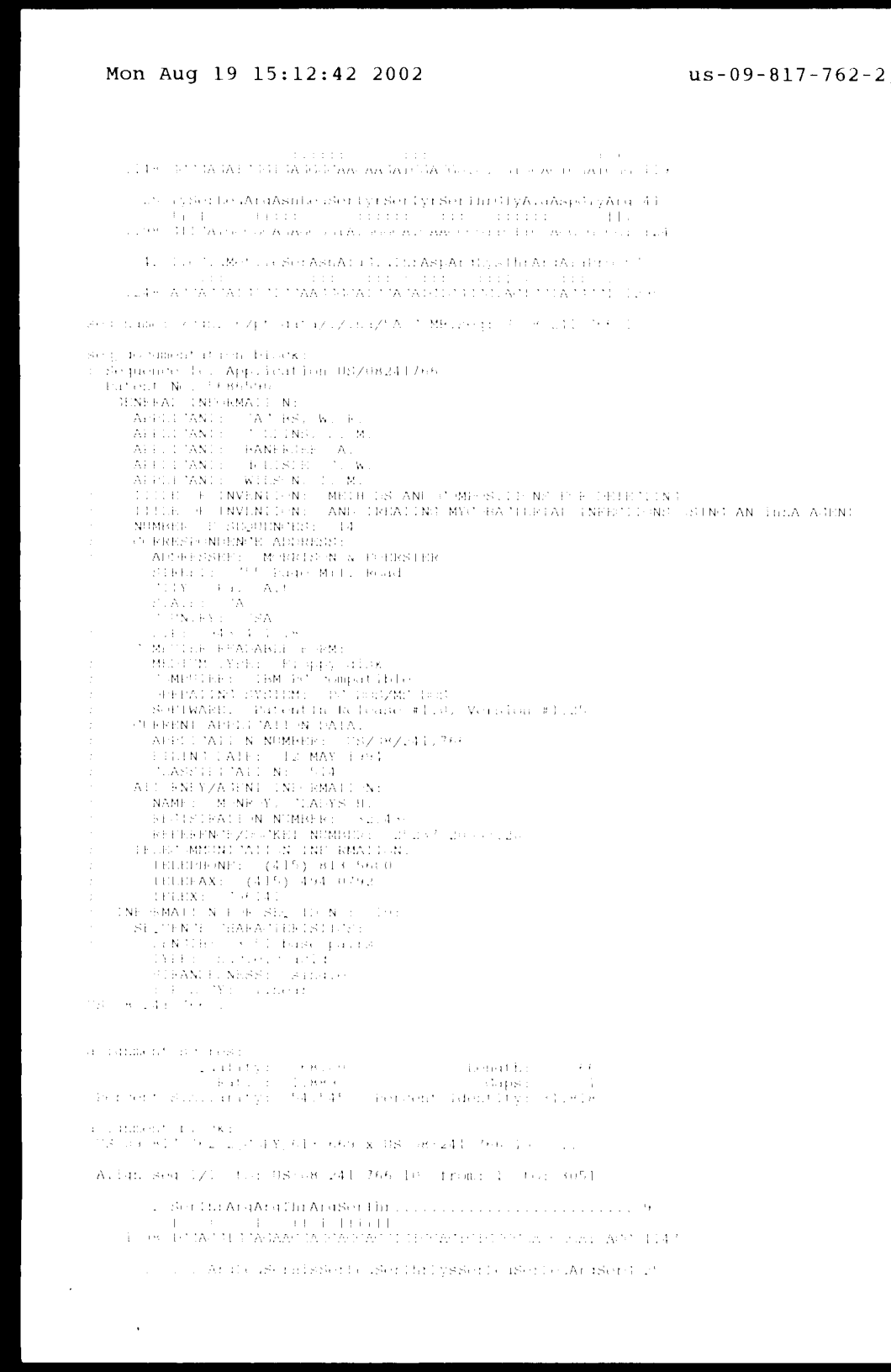
Alignment Block:

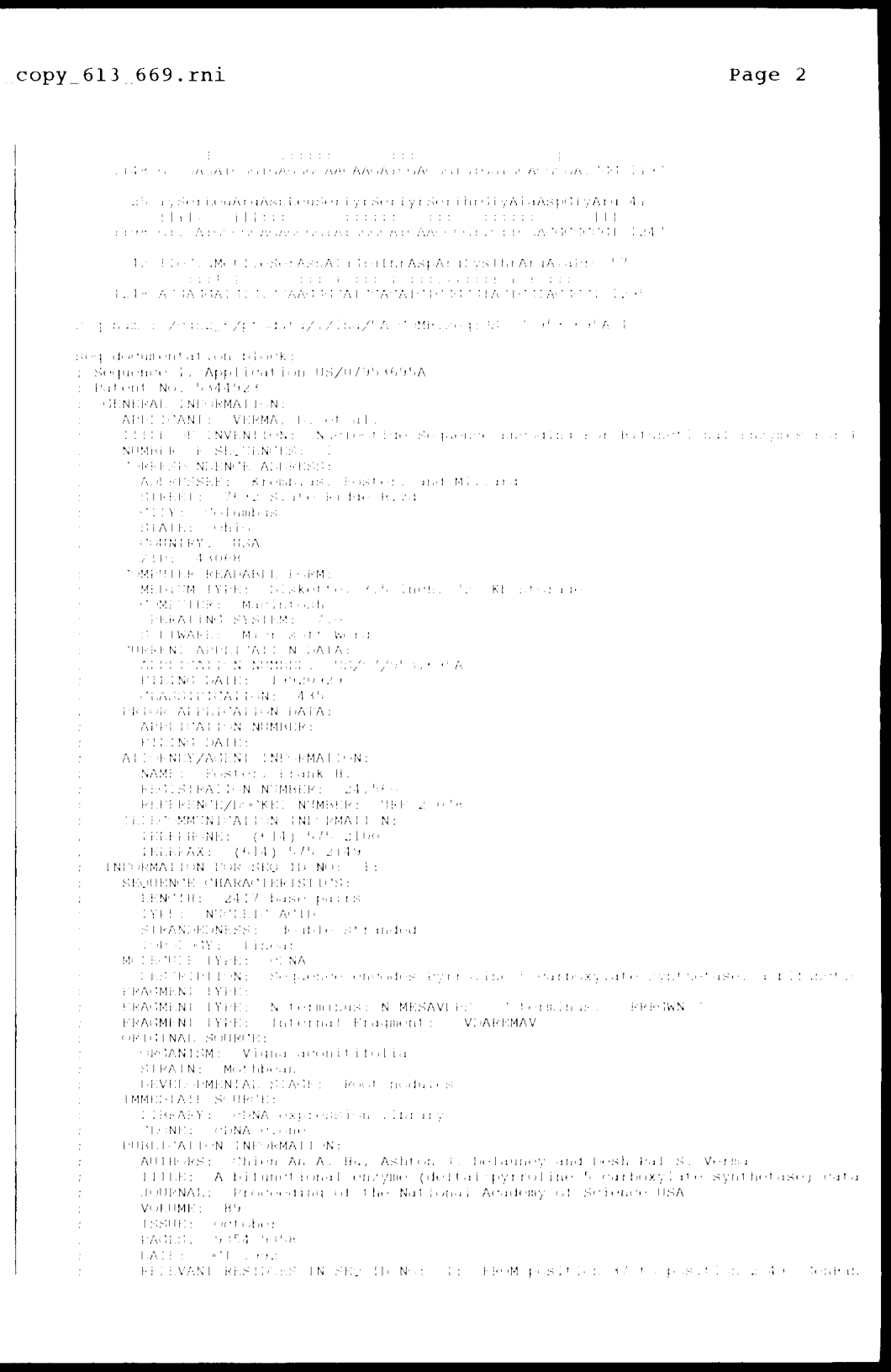
Alignment Block:

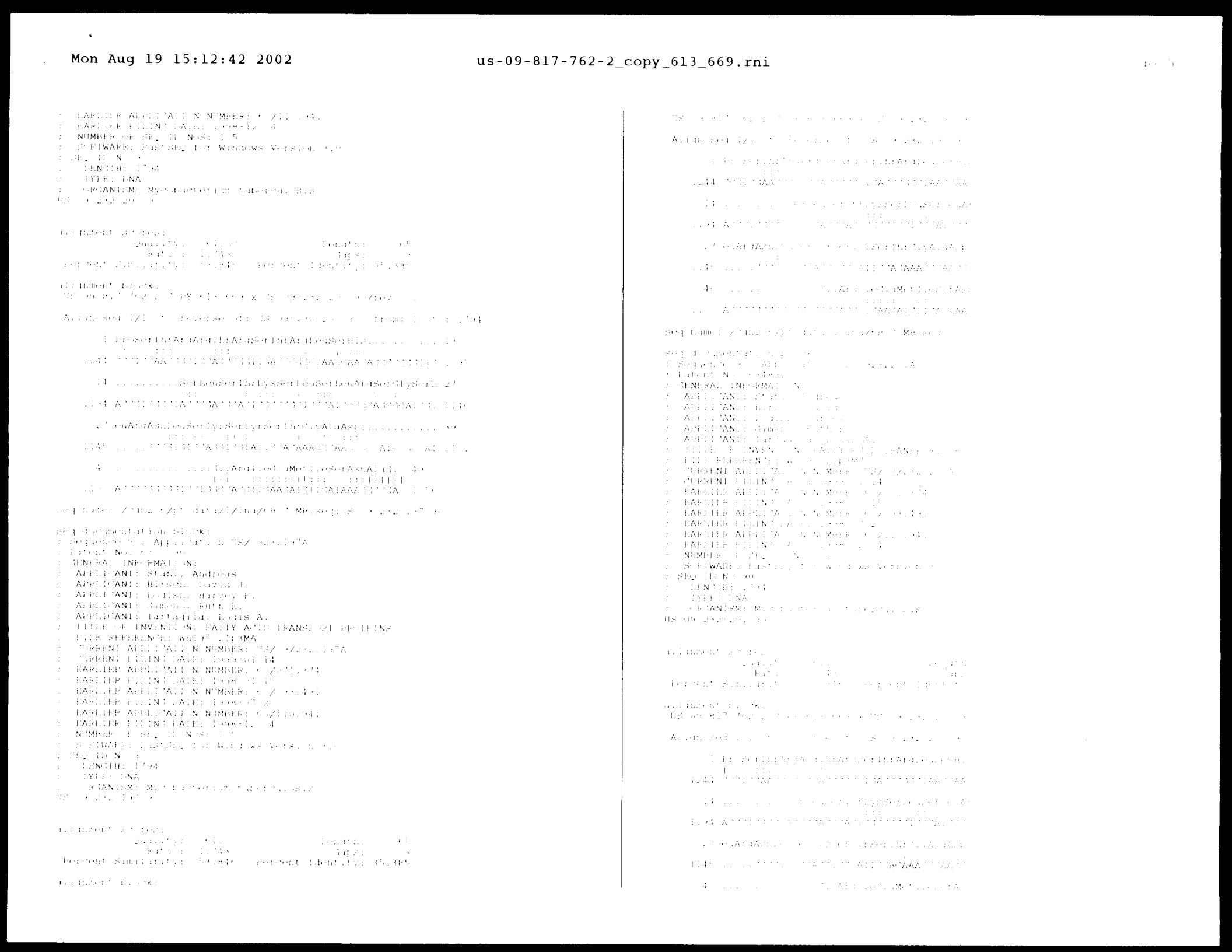
Sequence documentation block:
ID AA09841 standard; cDNA; 2417 BP.
XX
AC AA09841;
XX
DT 07-APR-1995 (first entry)
XX
DE Delta pyrroline 5-carboxylate synthetase (P5CS) coding sequence.
XX
KW Delta pyrroline 5-carboxylate synthetase; P5CS; proline; crop;
KW biosynthesis; metabolism; sodium chloride; salt; NaCl; drought; ds.
XX
OS Vigna aconitifolia.
XX
FH Key Location/qualifiers
FT CDS 47..2062
FT /*tag= a
FT /product= "delta pyrroline 5-carboxylate synthetase"
XX
RN [1]
XX
RP 1-SEP-1994.
XX
RP 2-SEP-1994; 92BS09569.
XX
RP 2-SEP-1994; 92US-09/569,695.
XX
RA (CSIS) UNIV CHILE STATE RES FOUND.
XX
RL Delauney A.J., Hu C.A., Verma D.P.S.
XX
DR WPI; 1994-285616/35.
DR P-PSDB; AAR61145.
XX
CC cDNA clone encoding bifunctional enzyme for proline produ
CC useful to prepare transgenic drought resistant plants
XX
CS Vigna 1; Figure 1; 17pp; English
XX
SQ Sequence 2417 BP; 669 A; 445 C; 582 G; 699 T; 0 other;

Alignment scores:
Quality: 64.00 Length: 54
Ratio: 1.778 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 47.037









```
[illegible]
[illegible]

Seq name: [illegible] US-09-232-191-22

Seq documentation block:
; Sequence 22, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p4ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,574
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,411
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 67
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
;   LENGTH: [illegible]
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-232-191-22

Alignment Scores:
          Quality:   61.00            Length:    65
          Ratio:   1.743              Gaps:     3
Percent Similarity: 53.846   Percent Identity: 41.538

Alignment Block:
US-09-817-762-2_COPY_613_669 x US-09-232-191-22/rev [illegible]

Align seq 1/1 to reverse of: US-09-232-191-22 from: 1 to: 2097

[illegible]

Seq name: [illegible] US-09-232-200-22

Seq documentation block:
; Sequence 22, Application US/09232200A
; Patent No. [illegible]
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p4MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,574
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,411
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
;   LENGTH: 2097
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-232-200-22

Alignment Scores:
          Quality:   61.00            Length:    65
          Ratio:   1.743              Gaps:     3
Percent Similarity: 53.846   Percent Identity: 41.538

Alignment Block:
US-09-817-762-2_COPY_613_669 x US-09-232-200-22/rev [illegible]

Align seq 1/1 to reverse of: US-09-232-200-22 from: 1 to: 2097

  1 [illegible]SerThrArgArgThrArgSer[illegible]ArgLeuSerHis.............. 14
[illegible]
1296 [illegible] 1241

 14 ..........SerLeuSerThrLysSerLeuSerLeuArgSerGlySerL 27
[illegible]
1251 [illegible] 1242

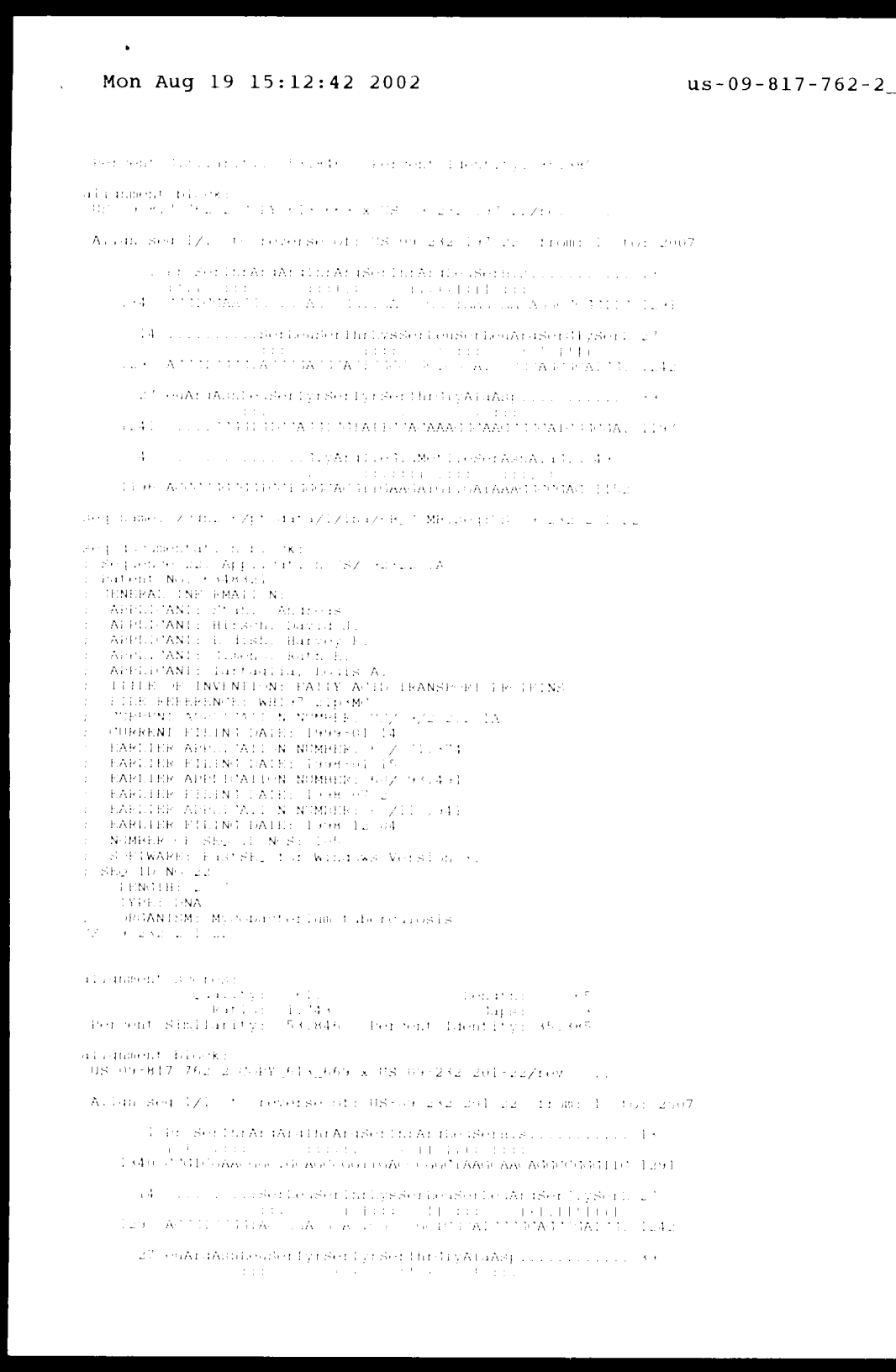
 27 euArgAsnLeuSerTyrSerTyrSerThrGlyAlaAsp............. 39
[illegible]
1241 [illegible] 1197

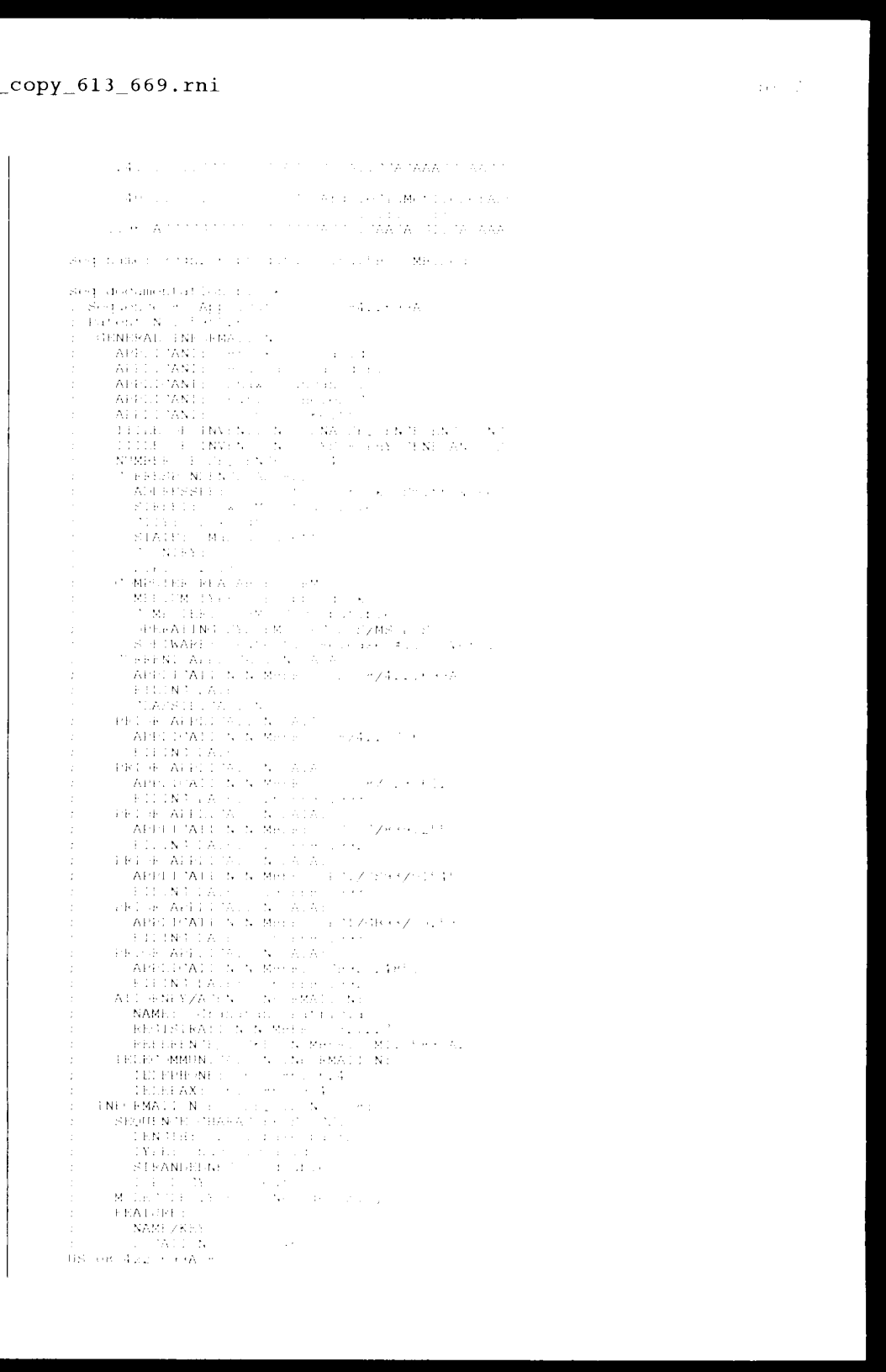
  4 ..................GlyArgIleGluMetIleSerAsnAlaGln 49
[illegible]
1196 [illegible] 1152

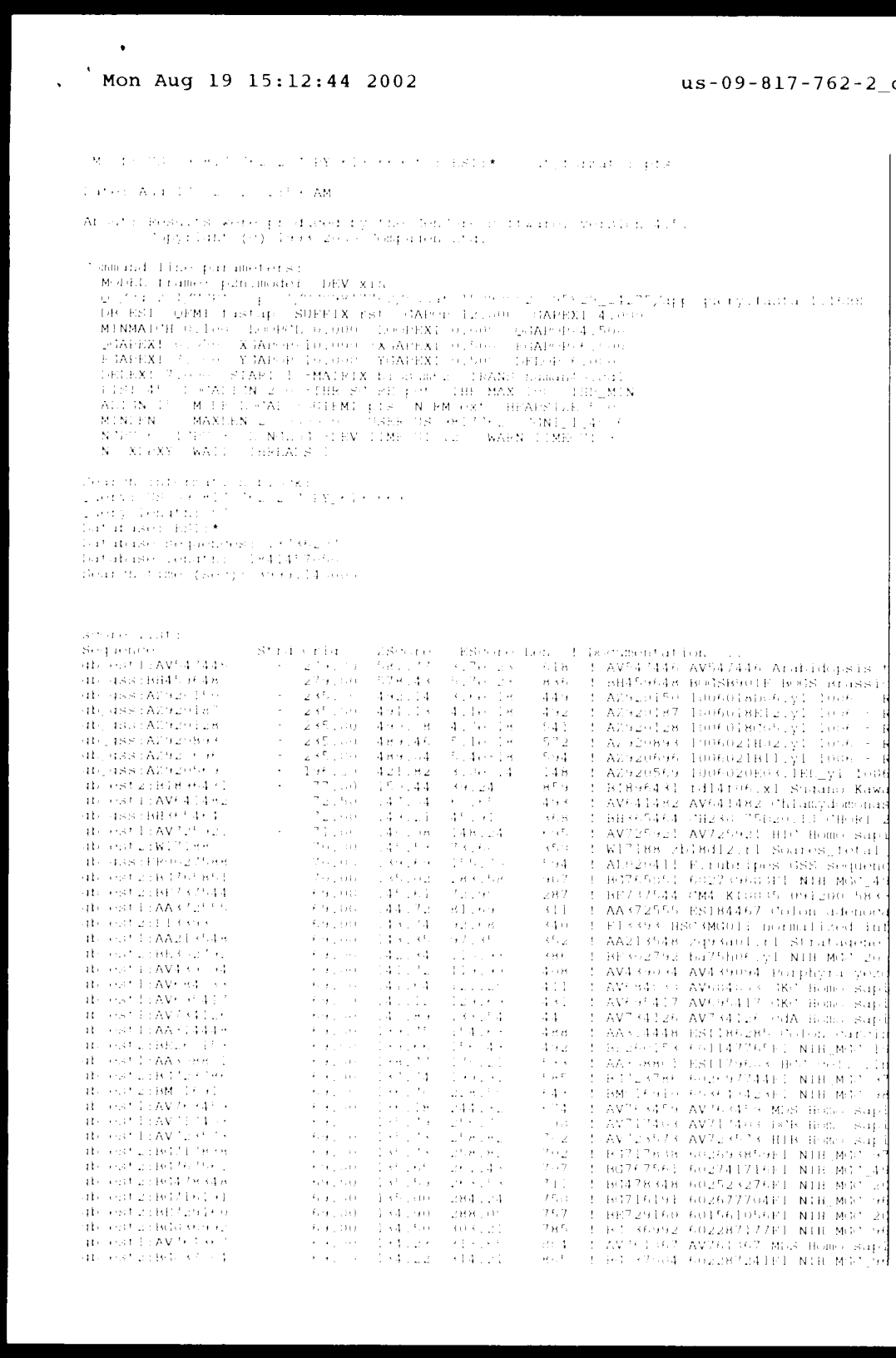
Seq name: [illegible] US-09-232-197-22

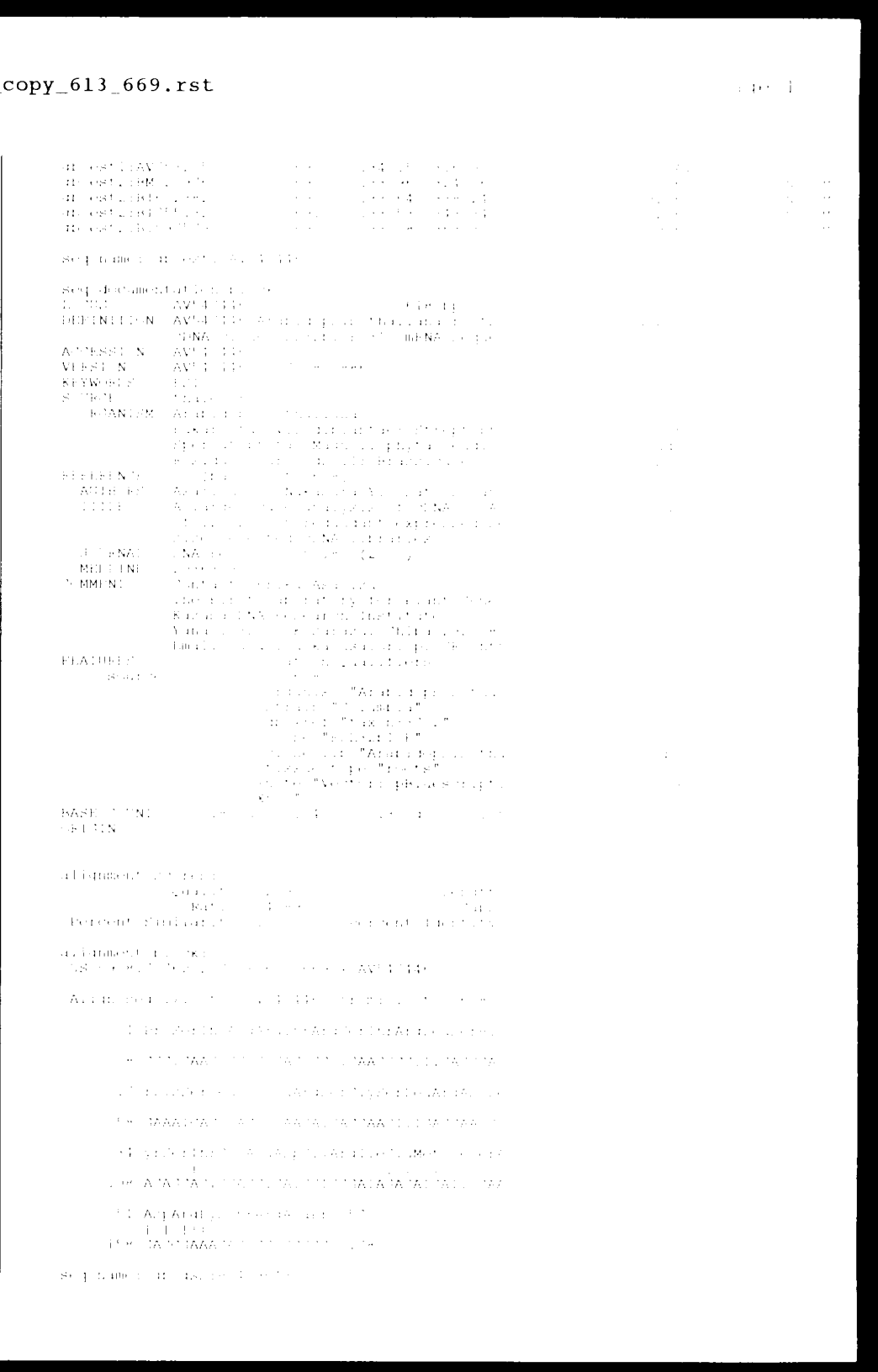
Seq documentation block:
; Sequence 22, Application US/09232197A
; Patent No. [illegible]
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p4MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,574
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,411
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
;   LENGTH: 2097
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-232-197-22

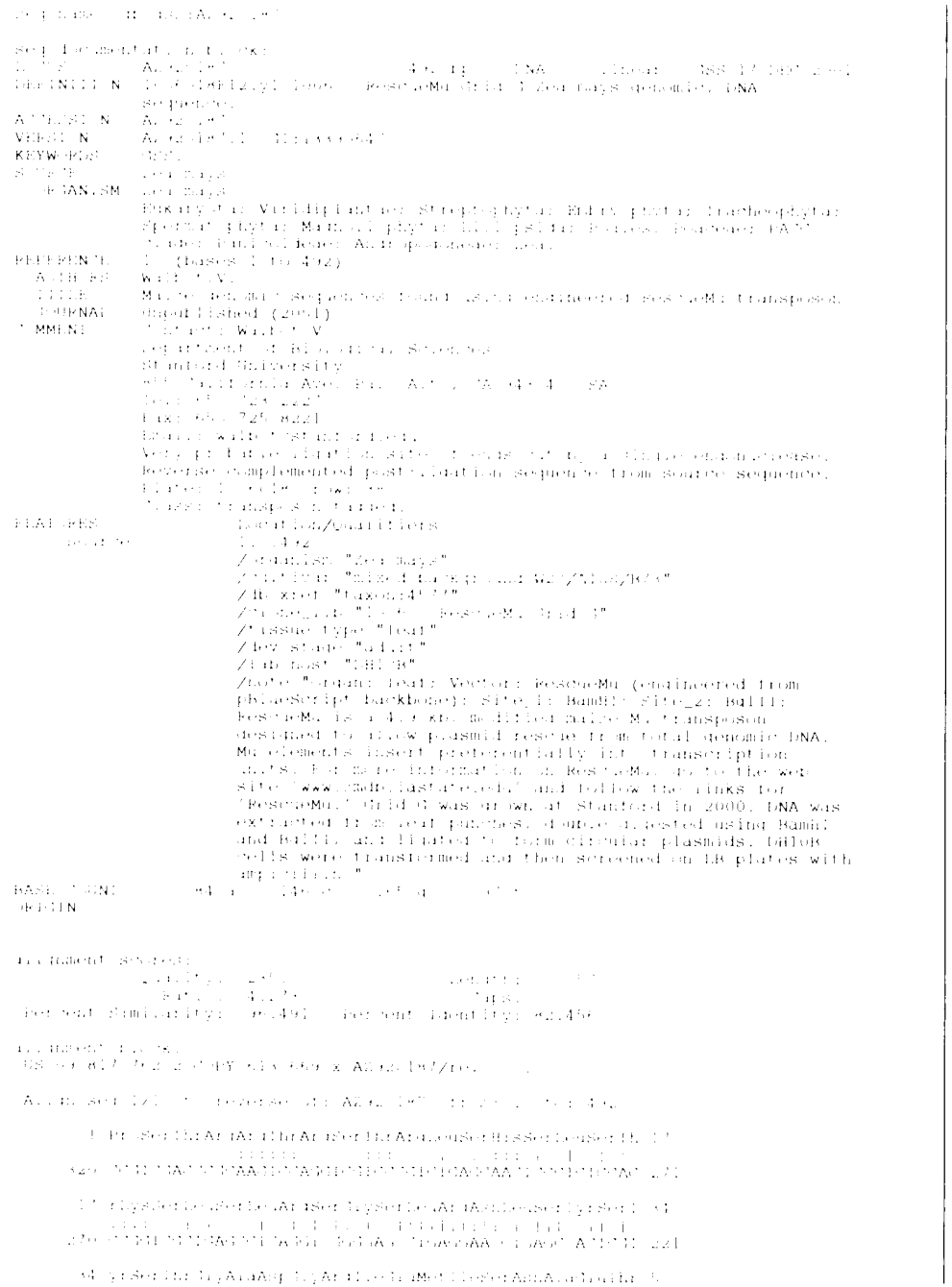
Alignment Scores:
          Quality:   61.00            Length:    65
          Ratio:   1.743              Gaps:     3
```



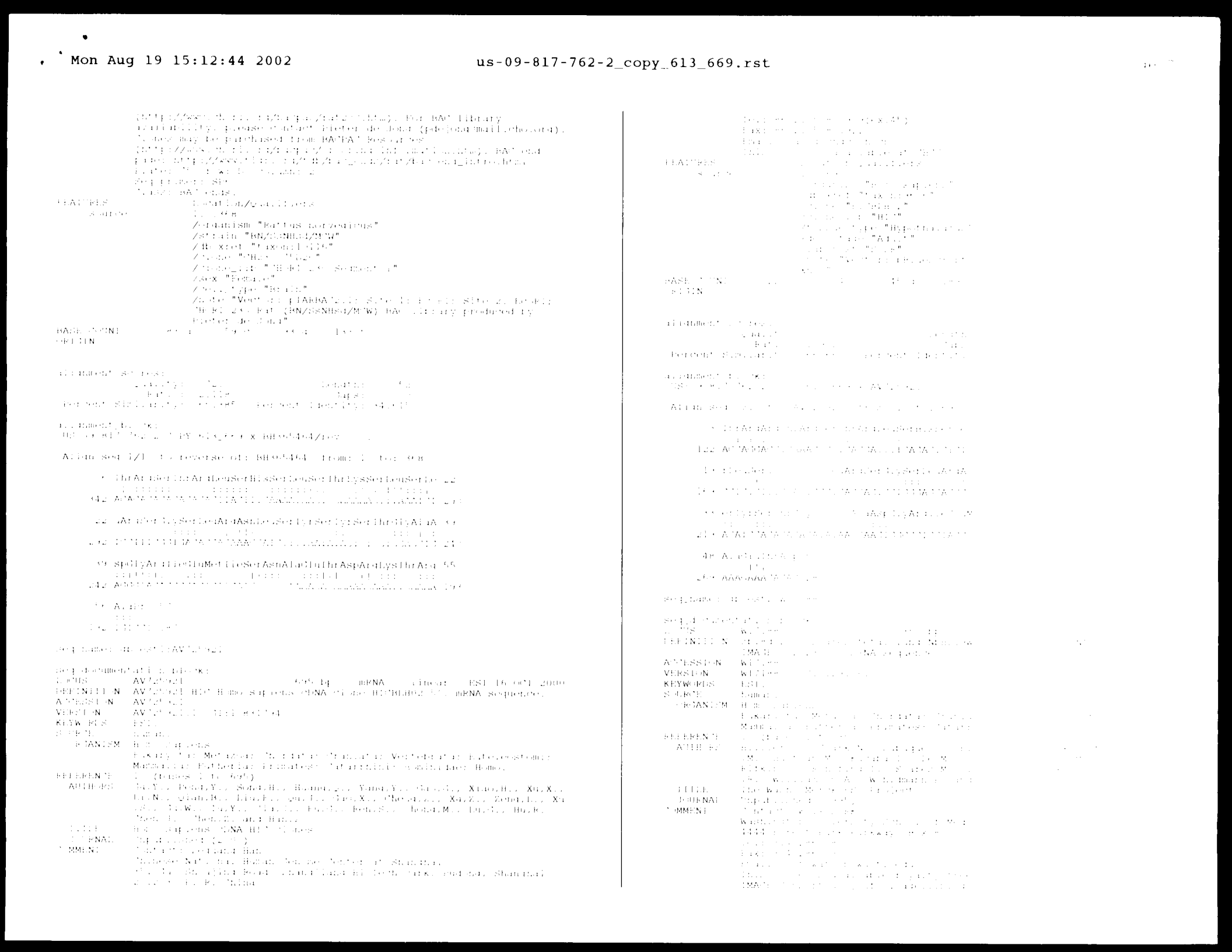


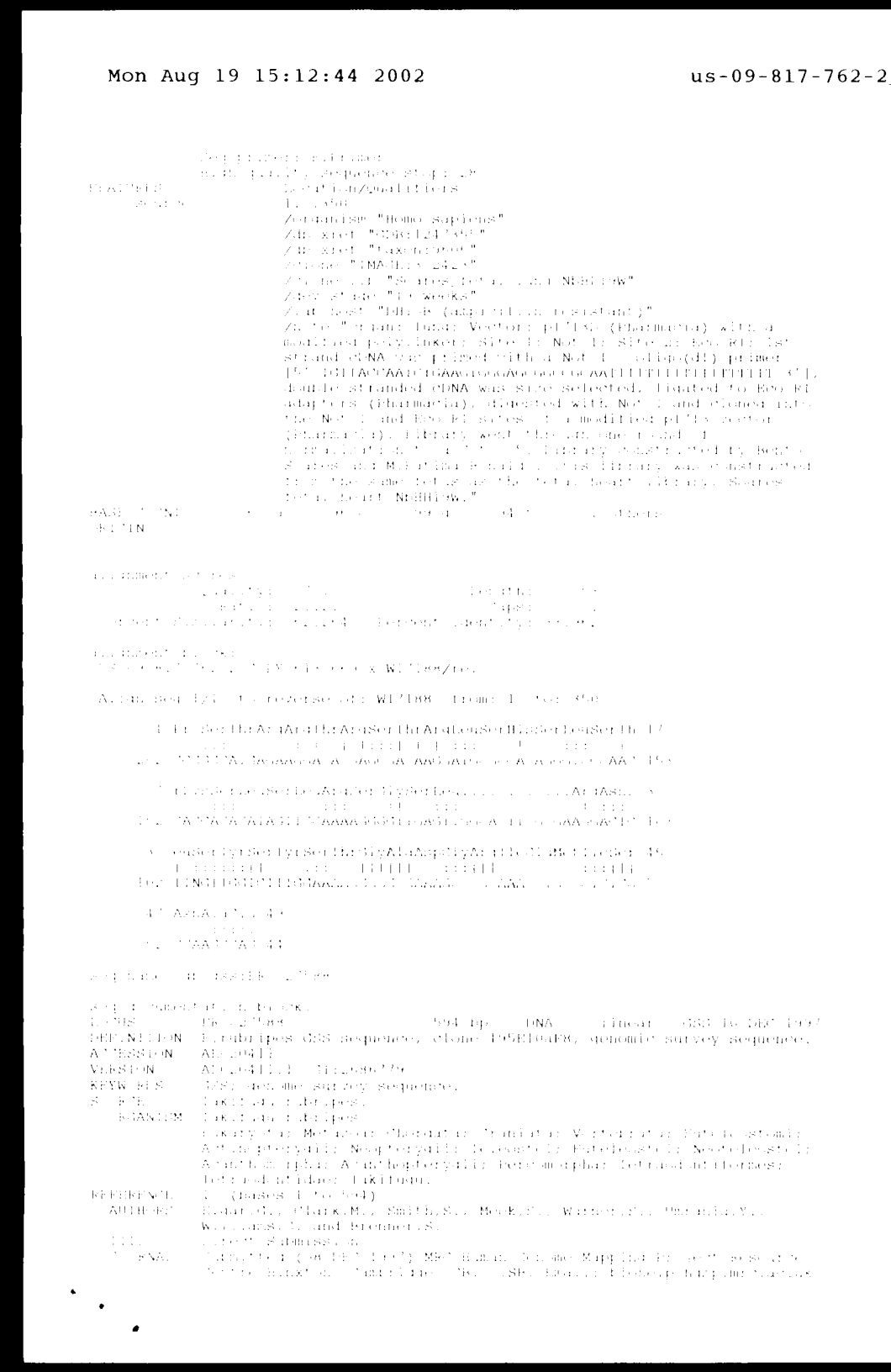


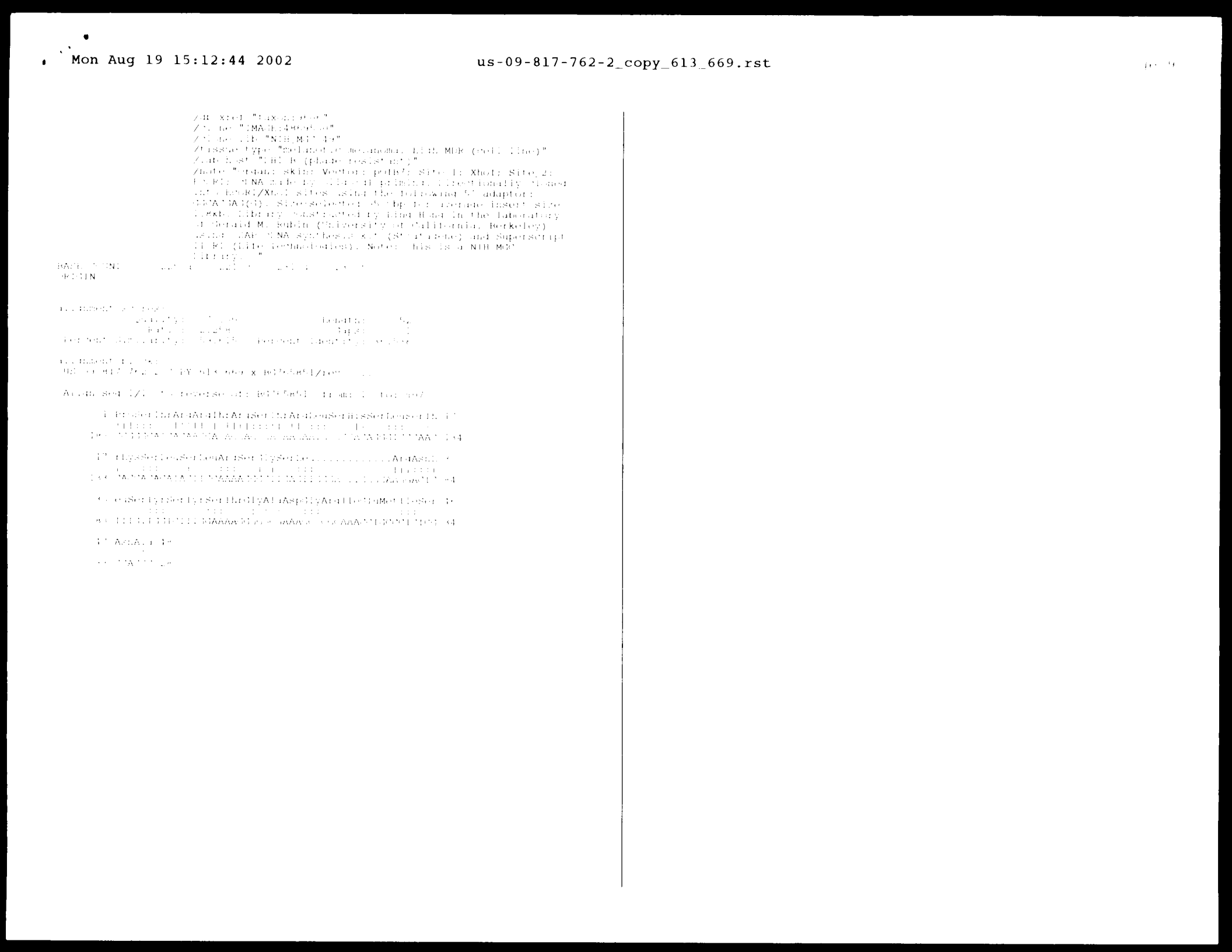


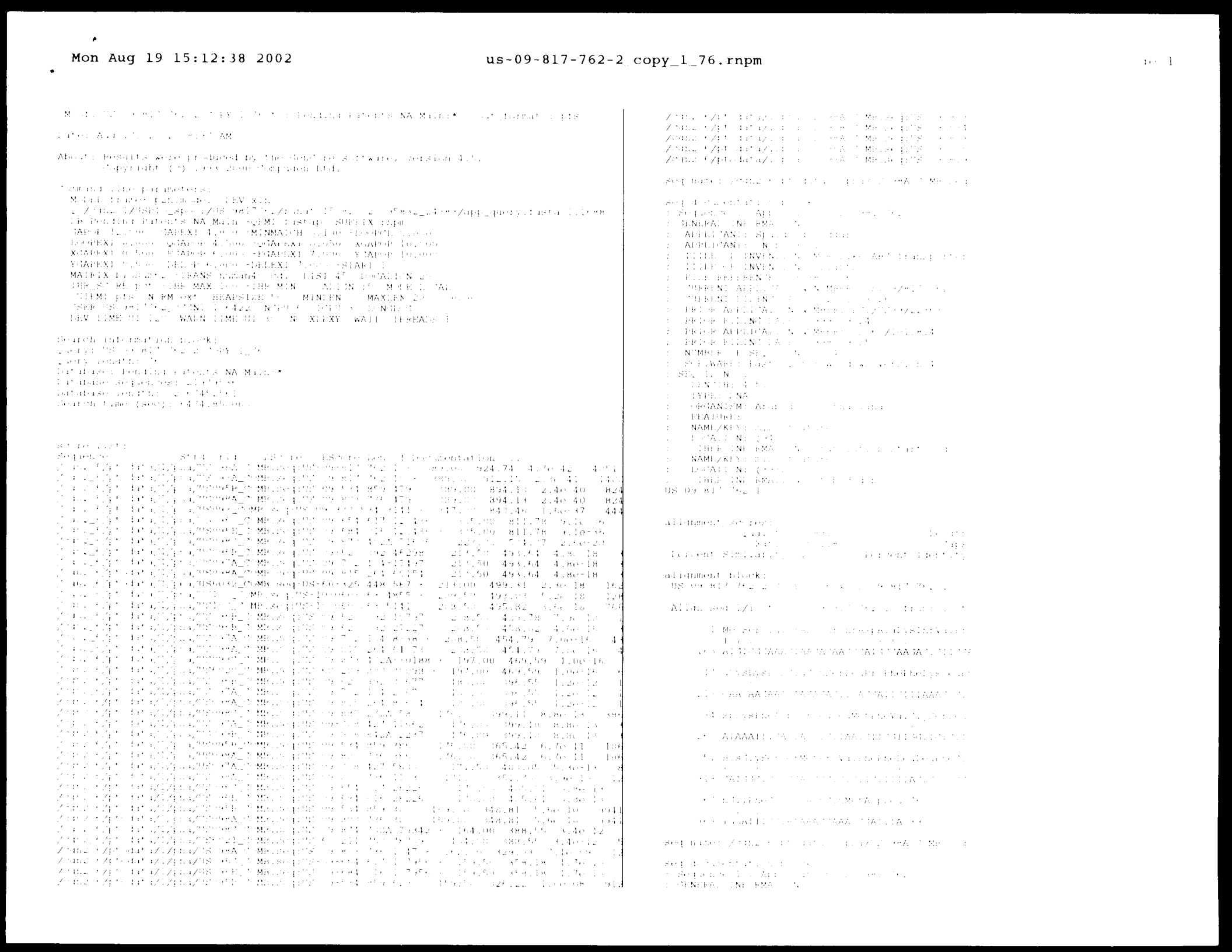


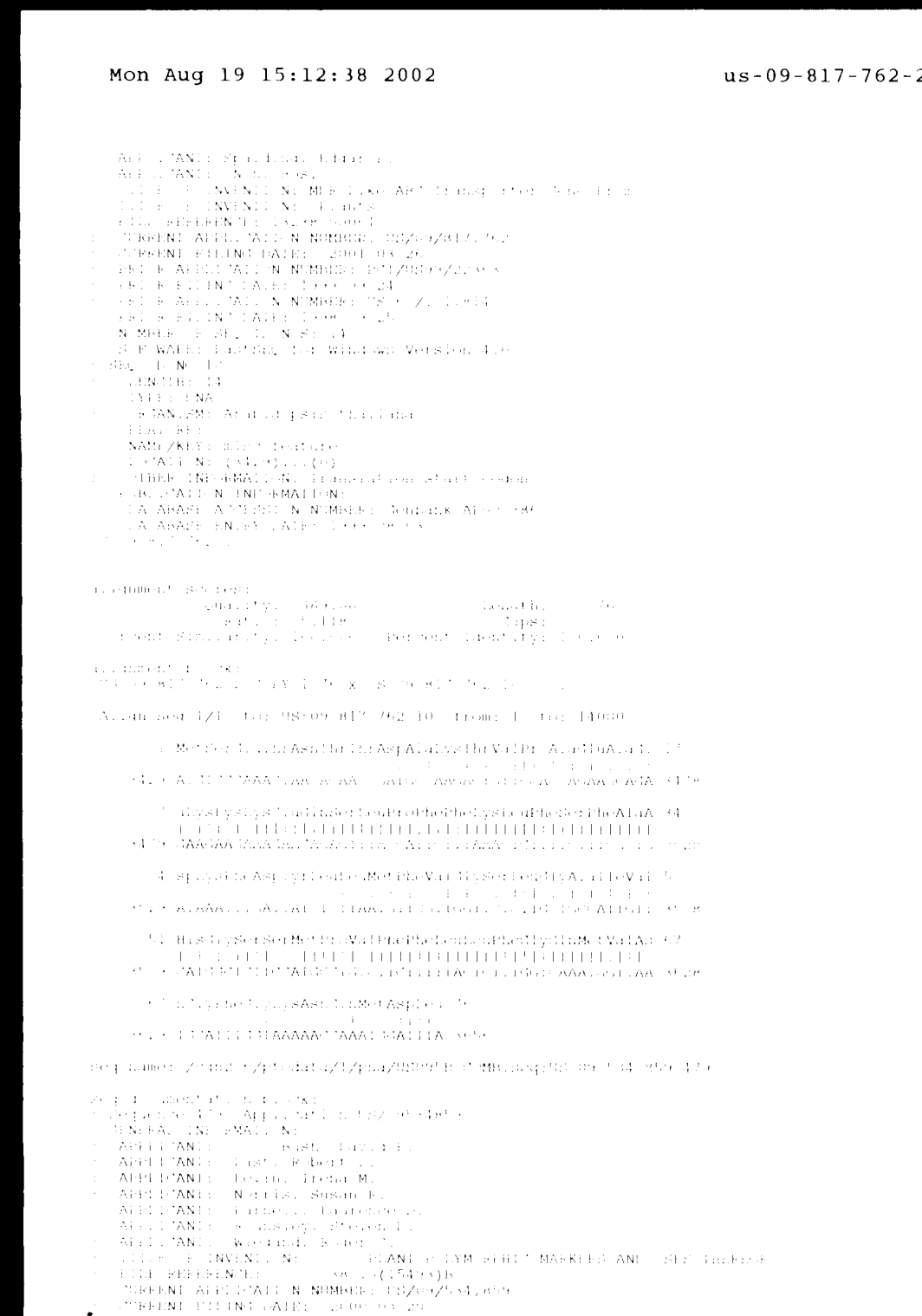
LOCUS AC262000 148 bp DNA linear SSS 17-DEC-2001
DEFINITION Zea mays Mu9 transposon, complete sequence.
ACCESSION AC262000
VERSION AC262000.1 GI:23345454
KEYWORDS SSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 148)
AUTHORS Walbot,V.
TITLE Mu9 Accession Sequences Used Within Unstable Zea mays Mu Transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
Stanford, CA 94305-5020
Tel: 650 725 8621
Fax: 650 725 8221
Email: walbot@stanford.edu
Very probable integration site of ends but by a sterile endonuclease.
Possible unexpected post 11bp, 5 sequence like acceptor sequence.
Extra 2 bp at ends.
Class: transposon tagged.
FEATURES Location/Qualifiers
source 1..148
/organism "Zea mays"
/mol_type "genomic DNA"
/db_xref "taxon:4577"
/clone_lib "Mu9 Mu transposon"
/tissue_type "root"
/dev_stage "adult"
/cell_type "DNA"
/note " Plant Cell Vol 13 Mu transposon (Accepted 11 Oct
pipingScript backbone); Site 1: BamHI; Site 2: BglII;
Mu9 is 4.4 kb, modified Mu1 Mu transposon.
Isolated by transposon tagging of sterile mutant genomic DNA.
Mu elements insert preferentially into transcription
units. For more information about Mu9 Mu transposon
site see www.MNR.Stanford.edu and links to

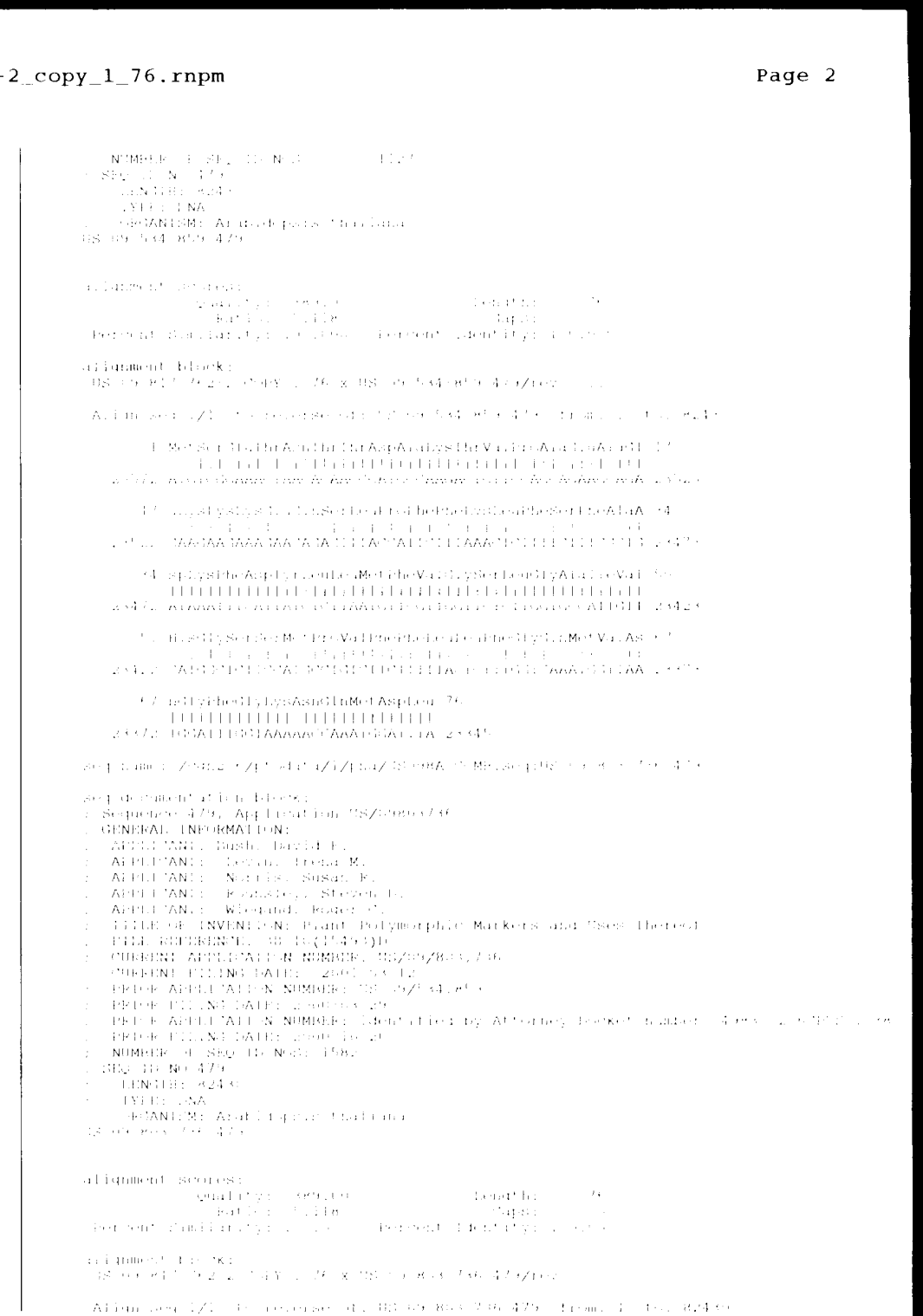


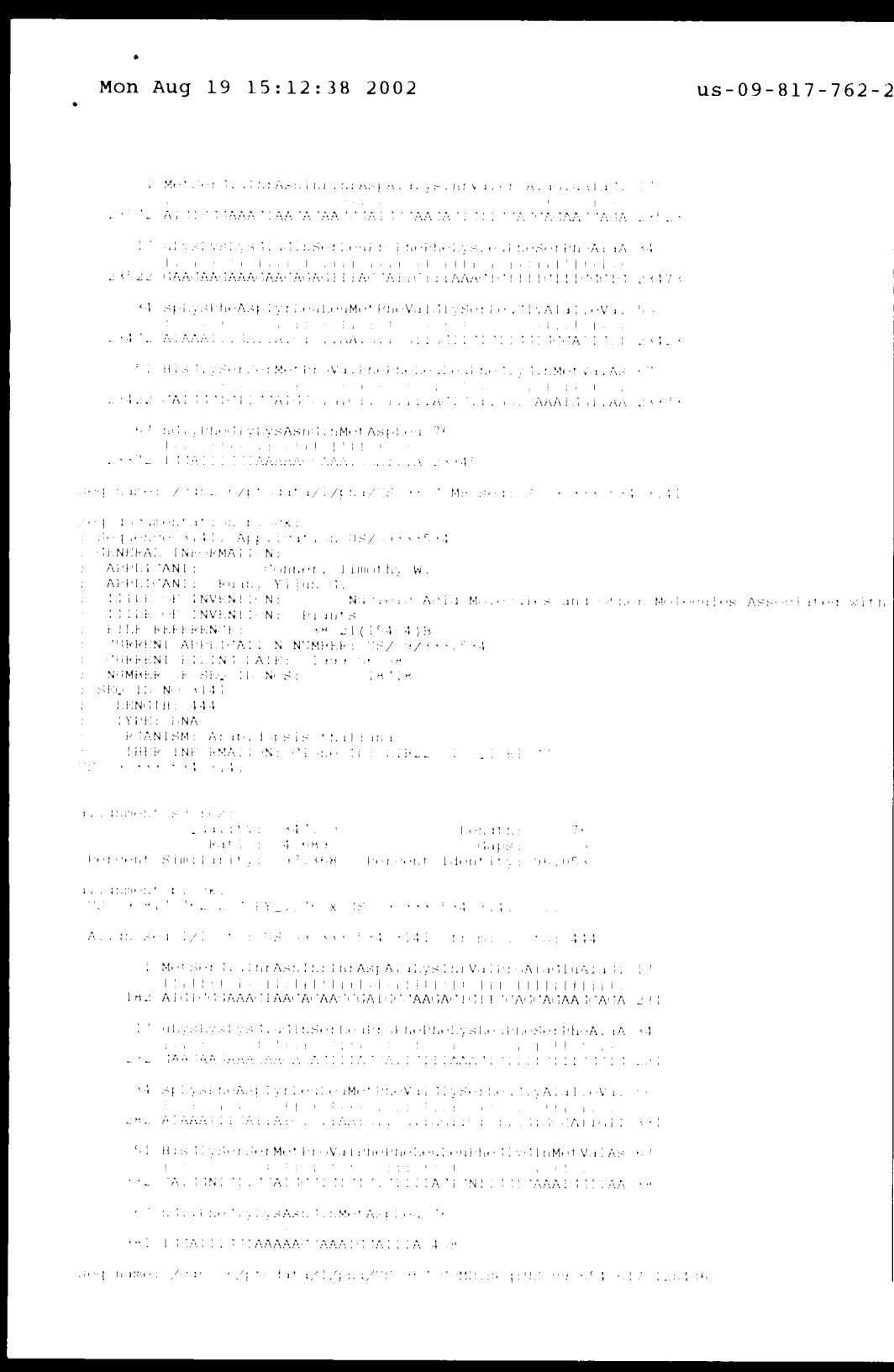


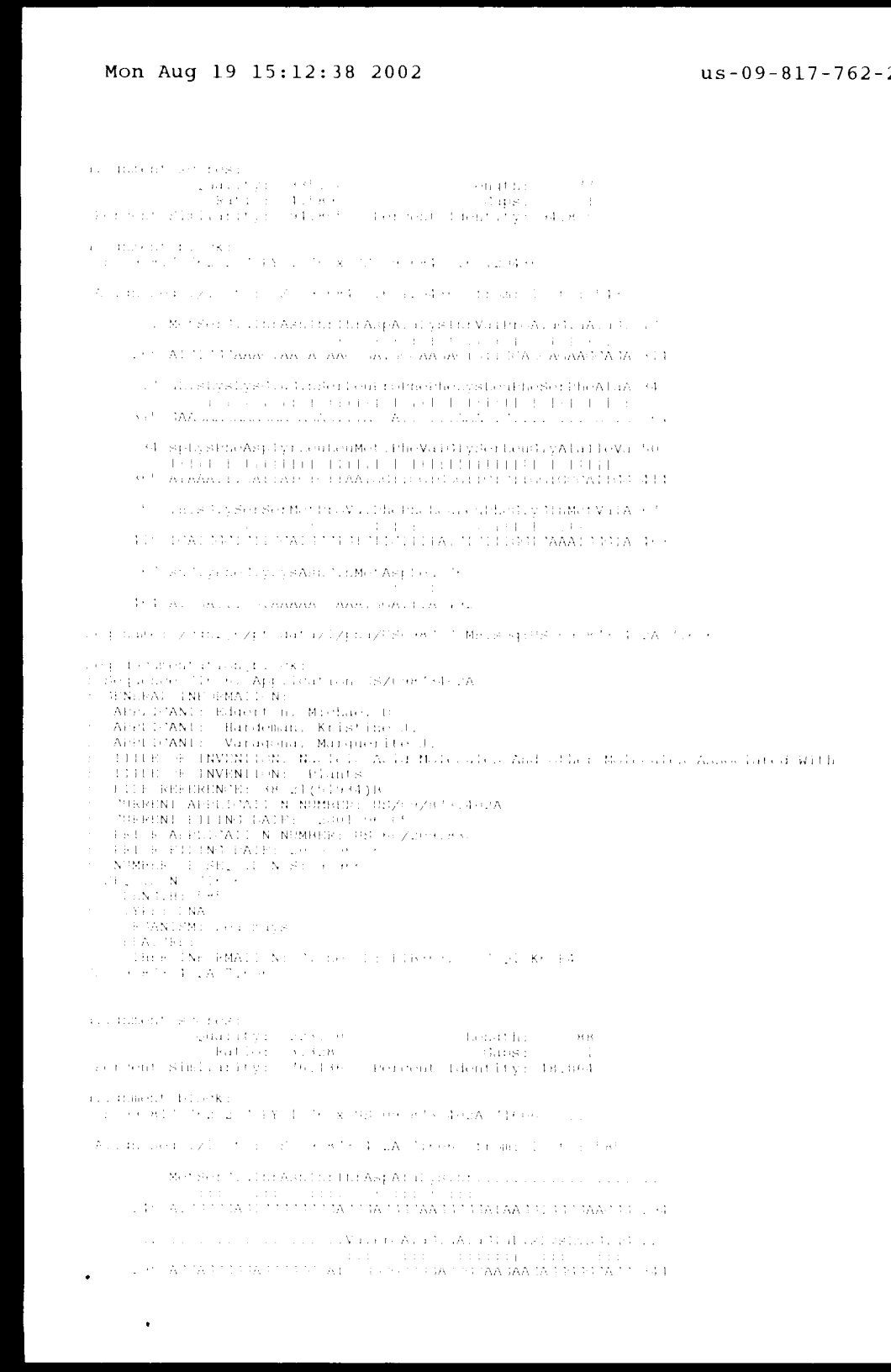














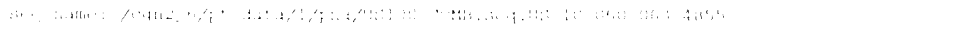








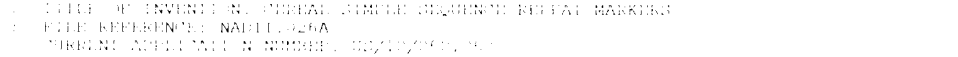




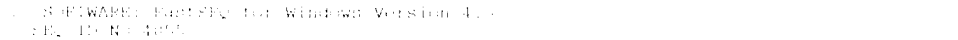












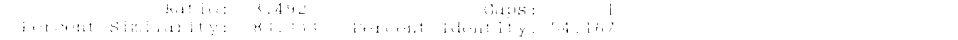


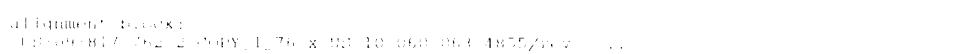


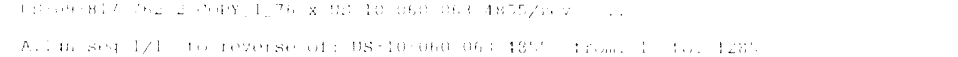


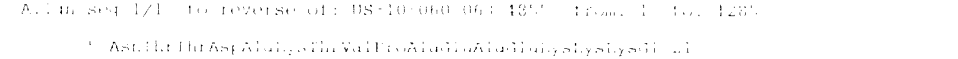


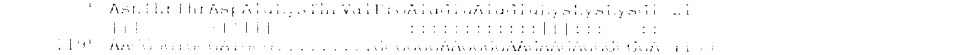


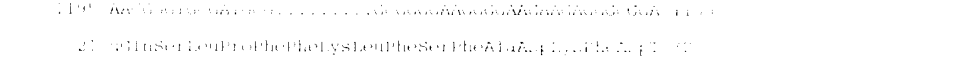


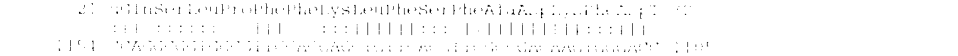


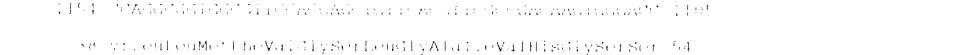


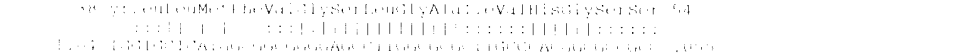


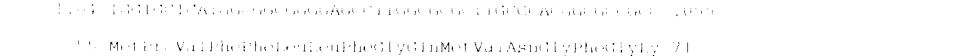


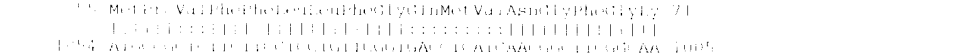






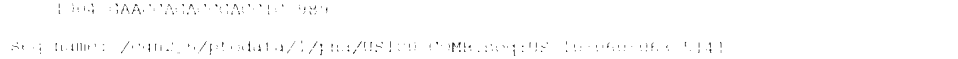


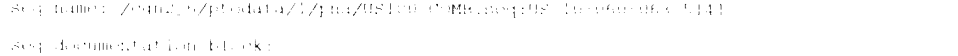










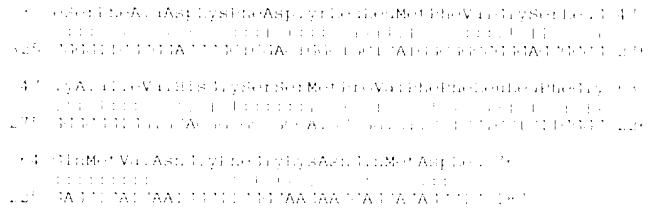


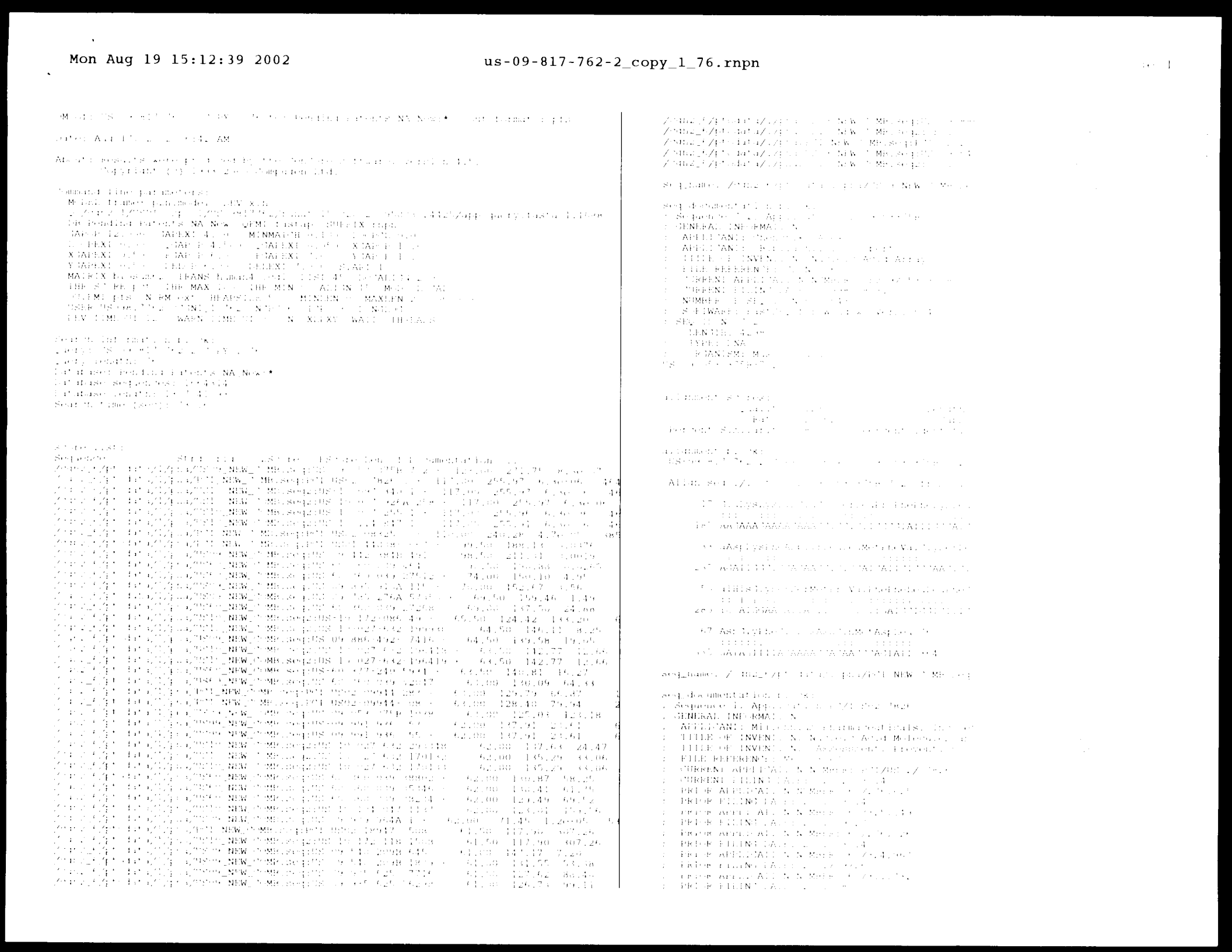










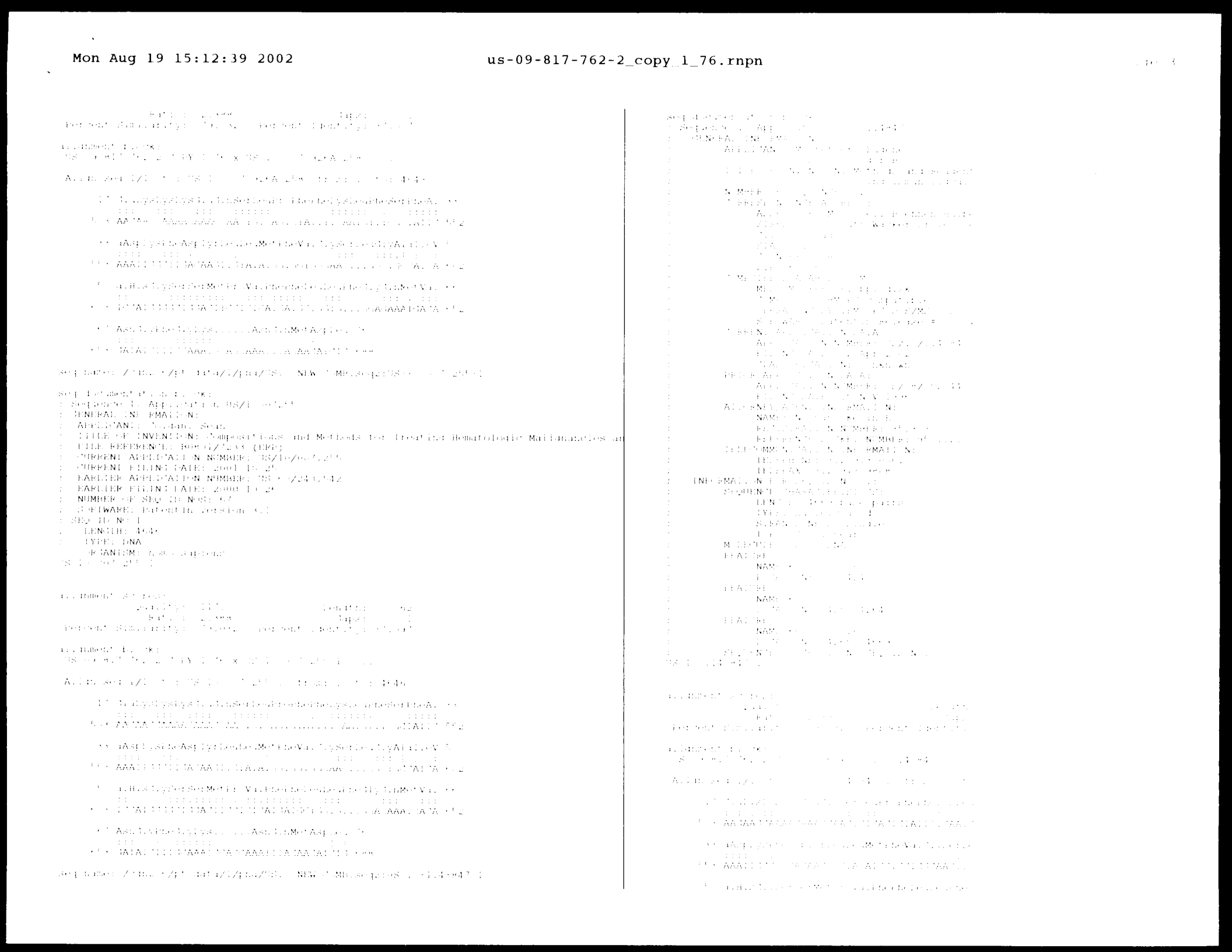


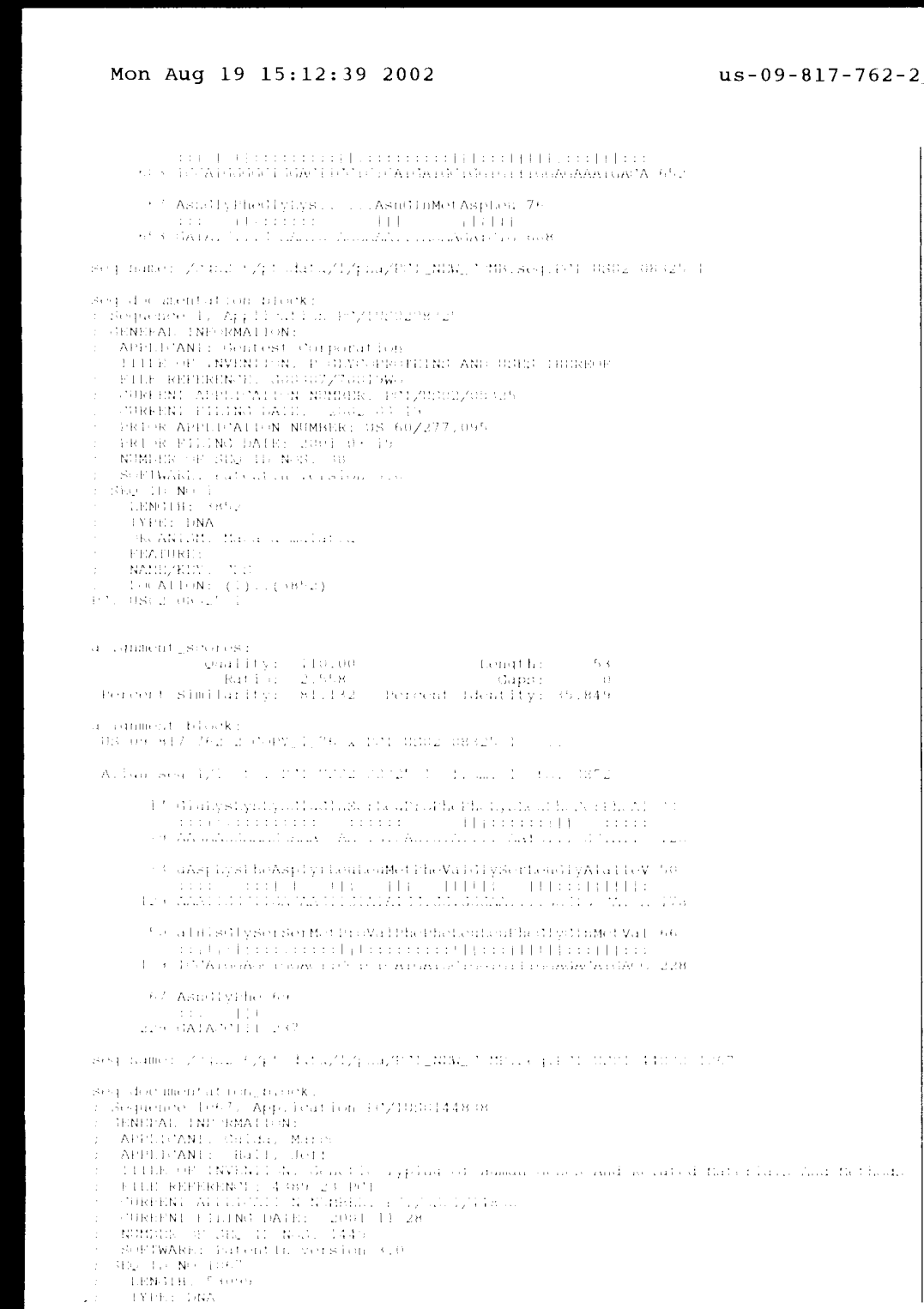
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 258
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: oligonucleotide cassette, 30th family 1c
; misc_feature(1)..(30)/tag), number 1 (Art001) primer
US-10-007-926A-258

; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 259
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: oligonucleotide cassette, 30th family 1c
; misc_feature(1)..(30)/tag), number 1 (Art001) primer
US-10-007-926A-259

; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 260
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: oligonucleotide cassette, 30th family 1c
; misc_feature(1)..(30)/tag), number 1 (Art001) primer
US-10-007-926A-260

; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 261
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: oligonucleotide cassette, 30th family 1c
; misc_feature(1)..(30)/tag), number 1 (Art001) primer
US-10-007-926A-261





; ORGANISM: Homo sapiens
PCT-US01 44838-1067

alignment scores:
Quality: 90.50 Length: 58
Ratio: 2.414 Gaps: 2
Percent Similarity: 74.138 Percent Identity: 41.379

alignment block:
US-09-817-762-2_COPY_1_76 x PCT-US01 44838-1067 .

Align seq 1/1 to PCT-US01 44838-1067 from 1 to 58399

Seq documentation block:
; Sequence 1, Application US/09817762
; GENERAL INFORMATION:
; APPLICANT: Genetech Corporation
; TITLE OF INVENTION: P-SELECTIN BINDING AND USES THEREOF
; FILE REFERENCE: 5020US/2(B07DW)
; CURRENT APPLICATION NUMBER: PCT/US2002/38525
; CURRENT FILING DATE: 2002-05-19
; PRIOR APPLICATION NUMBER: US 60/277,095
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3852
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3852)
PCT-US01 44838-1067

alignment scores:
Quality: 110.00 Length: 58
Ratio: 2.558 Gaps: 0
Percent Similarity: 81.132 Percent Identity: 45.849

alignment block:
US-09-817-762-2-COPY_1_76 x PCT-US01 44838-1067 .

Align seq 1/1 to PCT-US01 44838-1067 from 1 to 3852

Seq documentation block:
; Sequence 1, Application US/09817762
; GENERAL INFORMATION:
; APPLICANT: Genetech Corporation
; TITLE OF INVENTION: P-SELECTIN BINDING AND USES THEREOF
; FILE REFERENCE: 5020US/2(B07DW)
; CURRENT APPLICATION NUMBER: PCT/US2002/38525
; CURRENT FILING DATE: 2002-05-19
; PRIOR APPLICATION NUMBER: US 60/277,095
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3852
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3852)
PCT-US01 44838-1067

alignment scores:
Quality: 98.50 Length: 63
Ratio: 2.141 Gaps: 1
Percent Similarity: 75.410 Percent Identity: 41.148

alignment block:
US-09-817-762-2-COPY_1_76 x US-09-442-984B-491

Align seq 1/1 to US-09-442-984B-491 from 1 to 3924

Seq documentation block:
; Sequence 1, Application US/09817762
; GENERAL INFORMATION:
; APPLICANT: Childs, Mary
; APPLICANT: Ball, Jeff
; TITLE OF INVENTION: Recombinant cytokine receptor common gamma chain monoclonal antibodies and methods
; FILE REFERENCE: 4389.23 PCT
; CURRENT APPLICATION NUMBER: US/09/817,762
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1867
; LENGTH: 5000
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01 44838-1067

alignment scores:
Quality: 90.50 Length: 58
Ratio: 2.414 Gaps: 2
Percent Similarity: 74.138 Percent Identity: 41.379

alignment block:
US-09-817-762-2_COPY_1_76 x PCT-US01 44838-1067 .

Align seq 1/1 to PCT-US01 44838-1067 from 1 to 58399

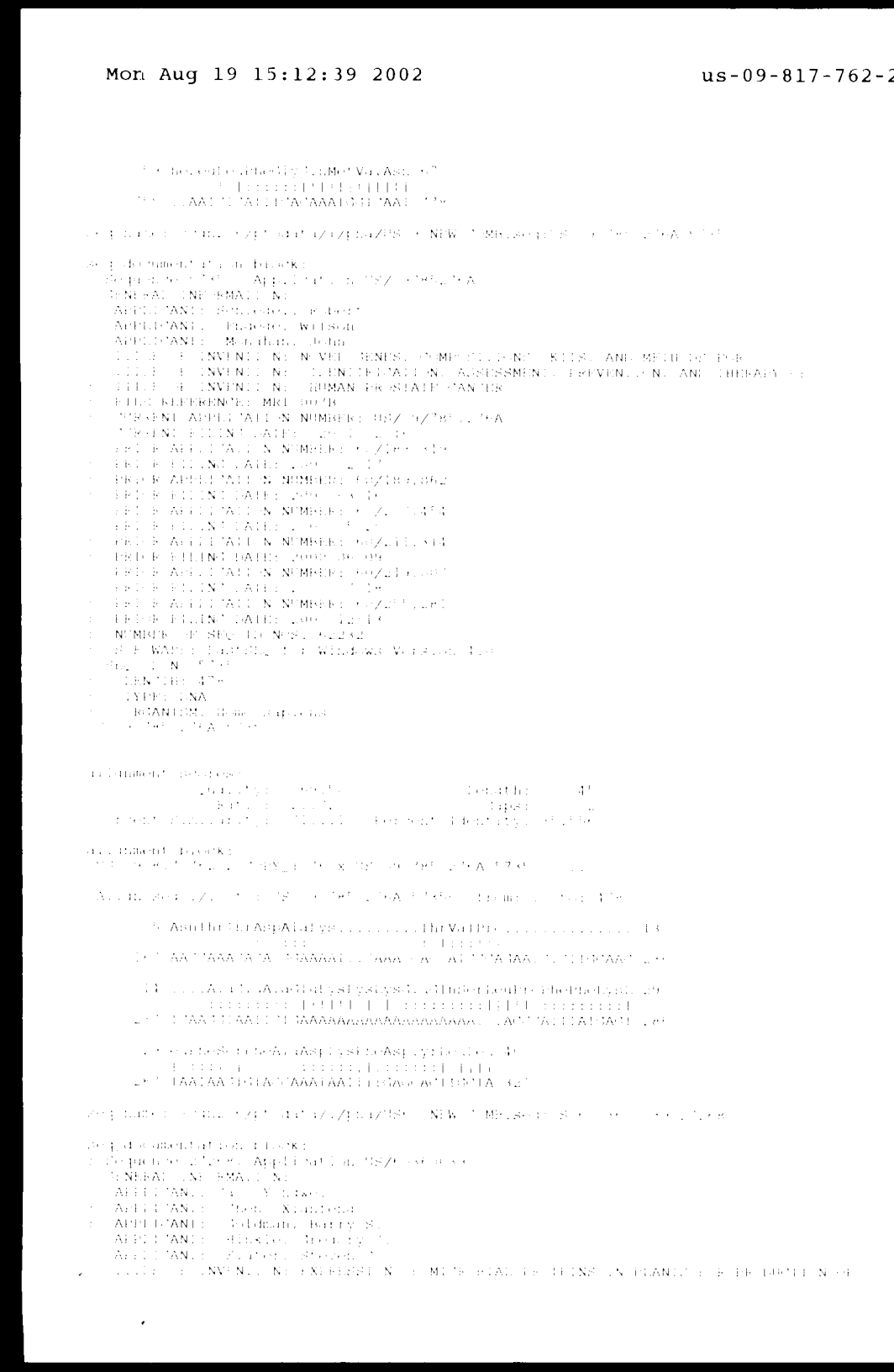
Seq documentation block:
; Sequence 491, Application US/09442384B
; GENERAL INFORMATION:
; APPLICANT: Chernchik, Alex
; APPLICANT: Takashev, Matvey
; TITLE OF INVENTION: Broad-Coverage Immunoassay Array
; FILE REFERENCE: CLON:006CIP15
; CURRENT APPLICATION NUMBER: US/09/442,384B
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 09/054,875
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 500
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491
; LENGTH: 3924
; TYPE: cDNA
; ORGANISM: Homo sapiens
US-09-442-984B-491

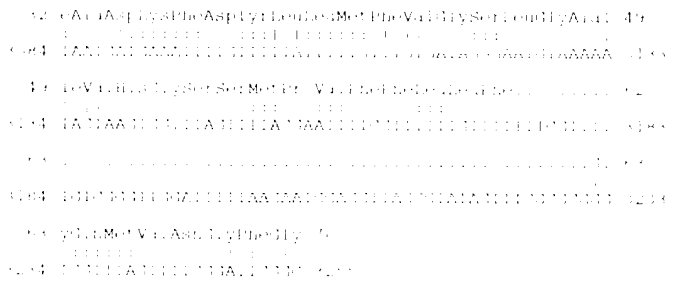
alignment scores:
Quality: 98.50 Length: 63
Ratio: 2.141 Gaps: 1
Percent Similarity: 75.410 Percent Identity: 41.148

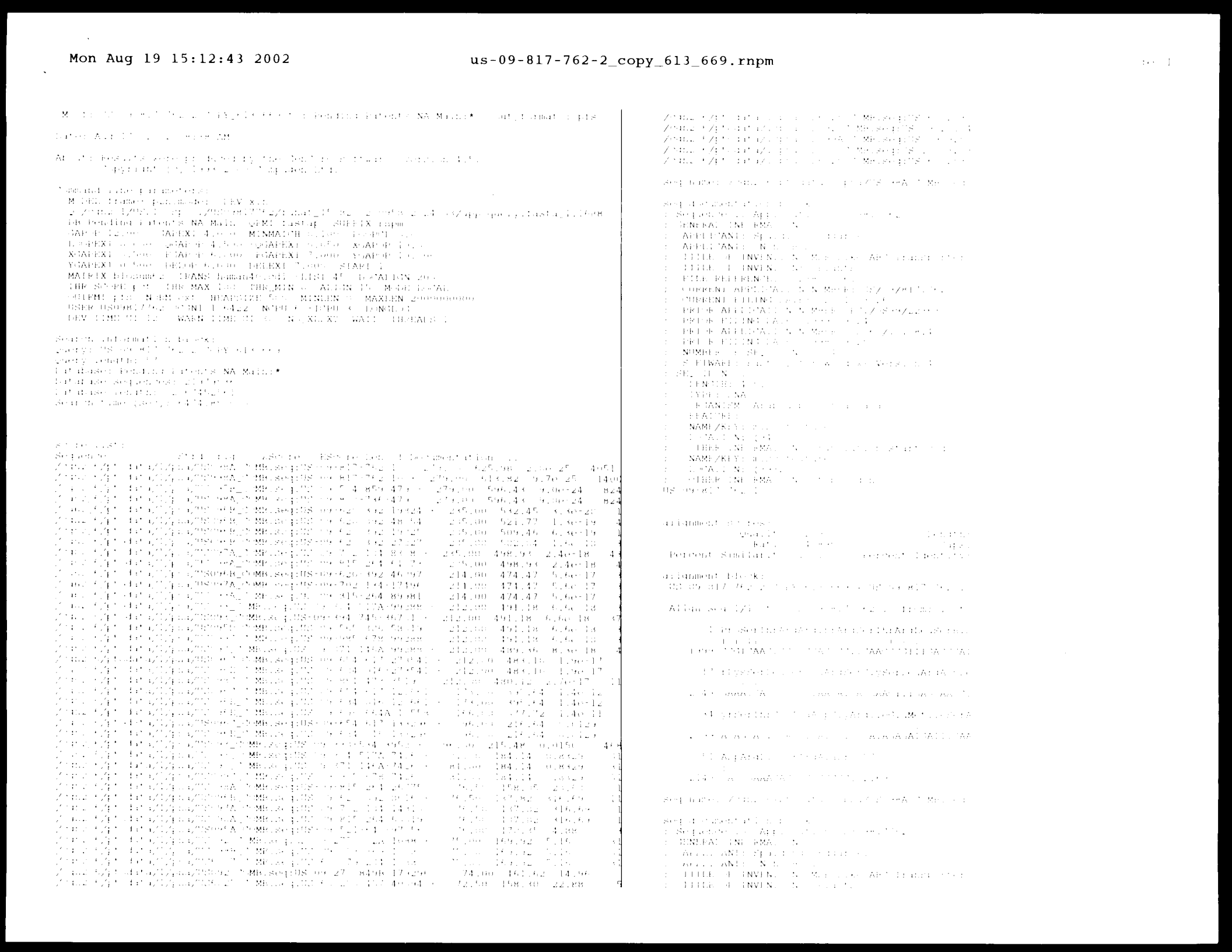
alignment block:
US-09-817-762-2-COPY_1_76 x US-09-442-984B-491

Align seq 1/1 to US-09-442-984B-491 from 1 to 3924

Seq documentation block:
; Sequence 491, Application US/09442384B
; GENERAL INFORMATION:
; APPLICANT: Chernchik, Alex
; APPLICANT: Takashev, Matvey
; TITLE OF INVENTION: Broad-Coverage Immunoassay Array
; FILE REFERENCE: CLON:006CIP15
; CURRENT APPLICATION NUMBER: US/09/442,384B
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 09/054,875
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 500
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491
; LENGTH: 3924
; TYPE: cDNA
; ORGANISM: Homo sapiens
US-09-442-984B-491







; FILE REFERENCE: 15238-0006;
; CURRENT APPLICATION NUMBER: US/09/817,762
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US/09/292,363
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: 08/940,811
; PRIOR FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 14000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5129)...(9)
; OTHER INFORMATION: n = A,T,C or G
; PUBLICATION INFORMATION:
; AUTHORS: Bush, David F.
; DATABASE ENTRY DATE: 1999-08-04
US-09-817-762-30

Alignment Scores:
Quality: 279.00 Length: 57
Ratio: 4.895 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment Block:
US-09-817-762-2_copy_613_669 x US-09-817-762-30

Align_Set 1/1 To reverse of: US-09-817-762-30 from: 1 to: 14000

1 ProSerThrArgArgThrArgSerThrArgLeuSerHisSerLeuSerTh 17
||||||||||||||||||||||||||||||||||||||||||||||||||||
7364 cAGGCCGAAAGACTCCTCGGCGCCT 7340

17 rLysSerLeuSerLeuArgSerGlySerLeuArgAsnLeuSerTyrSerT 34
|||||||||||||||||||||||||||||||||||||||||||||||||||
7363 cAAGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGA 7314

34 yrSerThrGlyAlaAspGlyArgProSerMetProSerAsnAlaGluThr 50
||||||||||||||||||||||||||||||||||||||||||||||||||
7313 ATGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGAT 7264

51 AspArgLysThrArgAlaPro 57
|||||||||||||||||||||
7263 CAGCCGAAAGACTCCTCGGCGCCT 7240

seq_name: /cgn2_6/ptodata/1/pubpct/US09817762_2.SEQ.seq:US-09-817-762-2:1:1:669:1:gcg:3:US:479:479

Seq_documentation_Block:
; Sequence 479, Application US/09534859
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Last, Robert L.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Purnell, Lawrence D.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: PLANT POLYMORPHIC MARKERS AND USES THEREOF
; FILE REFERENCE: 08-30(15479)B
; CURRENT APPLICATION NUMBER: US/09/534,859
; CURRENT FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 1127
; SEQ ID NO 479
; LENGTH: 82430
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-534-859-479

Alignment Scores:
Quality: 279.00 Length: 57
Ratio: 4.895 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment Block:
US-09-817-762-2_copy_613_669 x US-09-534-859-479/rev

Align_Set 1/1 To reverse of: US-09-534-859-479 from: 1 to: 82430

1 ProSerThrArgArgThrArgSerThrArgLeuSerHisSerLeuSerTh 17
||||||||||||||||||||||||||||||||||||||||||||||||||||
19667 cGGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGA 19716

17 rLysSerLeuSerLeuArgSerGlySerLeuArgAsnLeuSerTyrSerT 34
||||||||||||||||||||||||||||||||||||||||||||||||||||
19717 cAAGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGAT 19766

34 yrSerThrGlyAlaAspGlyArgProSerMetProSerAsnAlaGluThr 50
||||||||||||||||||||||||||||||||||||||||||||||||||
19767 ATGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGAT 19816

51 AspArgLysThrArgAlaPro 57
|||||||||||||||||||||
19817 CAGCCGAAAGACTCCTCGGCGCCT 19840

seq_name: /cgn2_6/ptodata/1/pubpct/US09817762_2.SEQ.seq:US-09-534-859-479:1:1:669:1:gcg:3:US:479:479

Seq_documentation_Block:
; Sequence 479, Application US/09534859
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 08-30(15479)D
; CURRENT APPLICATION NUMBER: US/09/534,859
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 09/292,363 (Identified by Attorney Docket number 08-30(15479)B)
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 479
; LENGTH: 82430
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-817-762-479

Alignment Scores:
Quality: 279.00 Length: 57
Ratio: 4.895 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

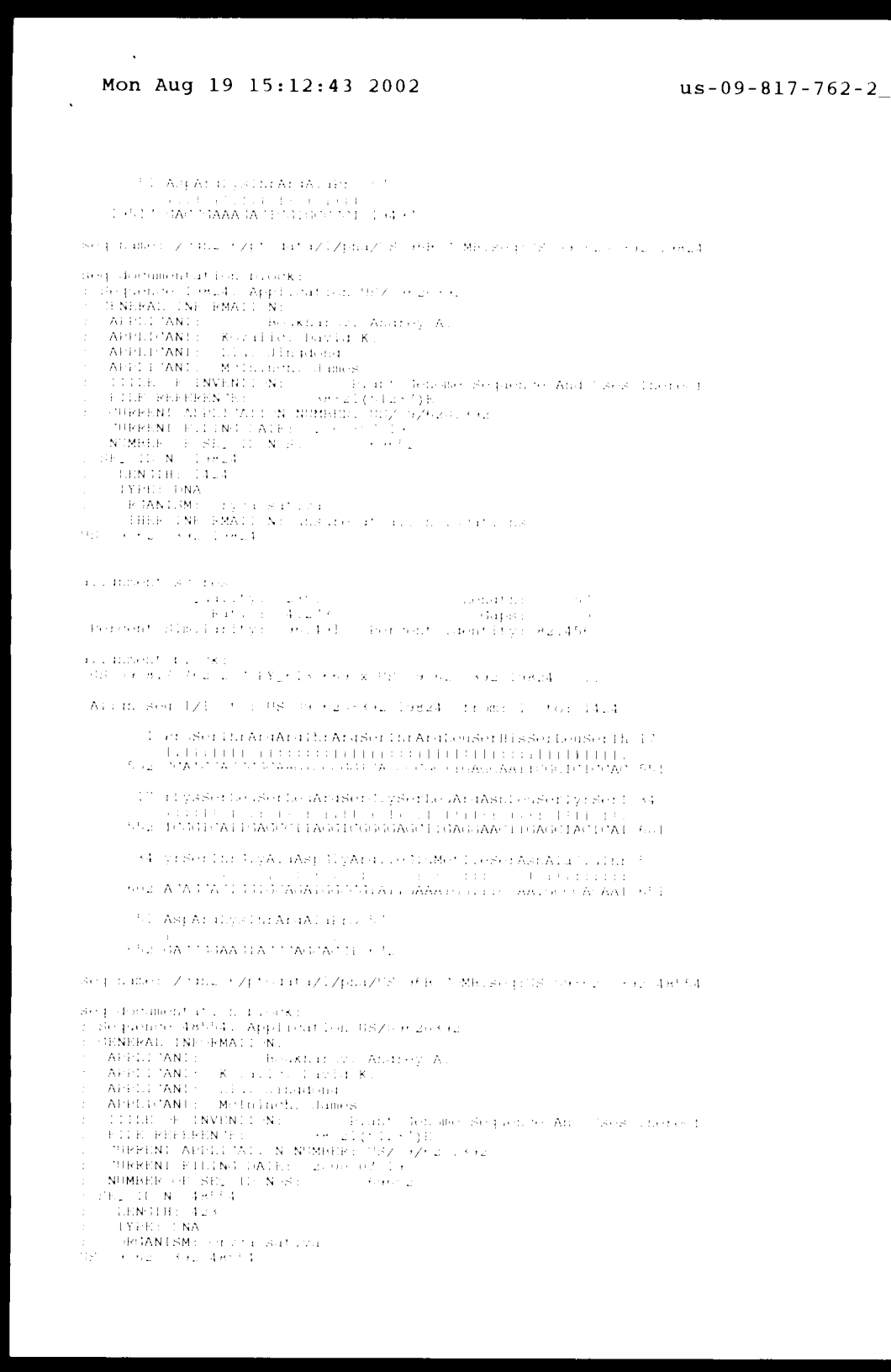
Alignment Block:
US-09-817-762-2_copy_613_669 x US-09-817-762-479/rev

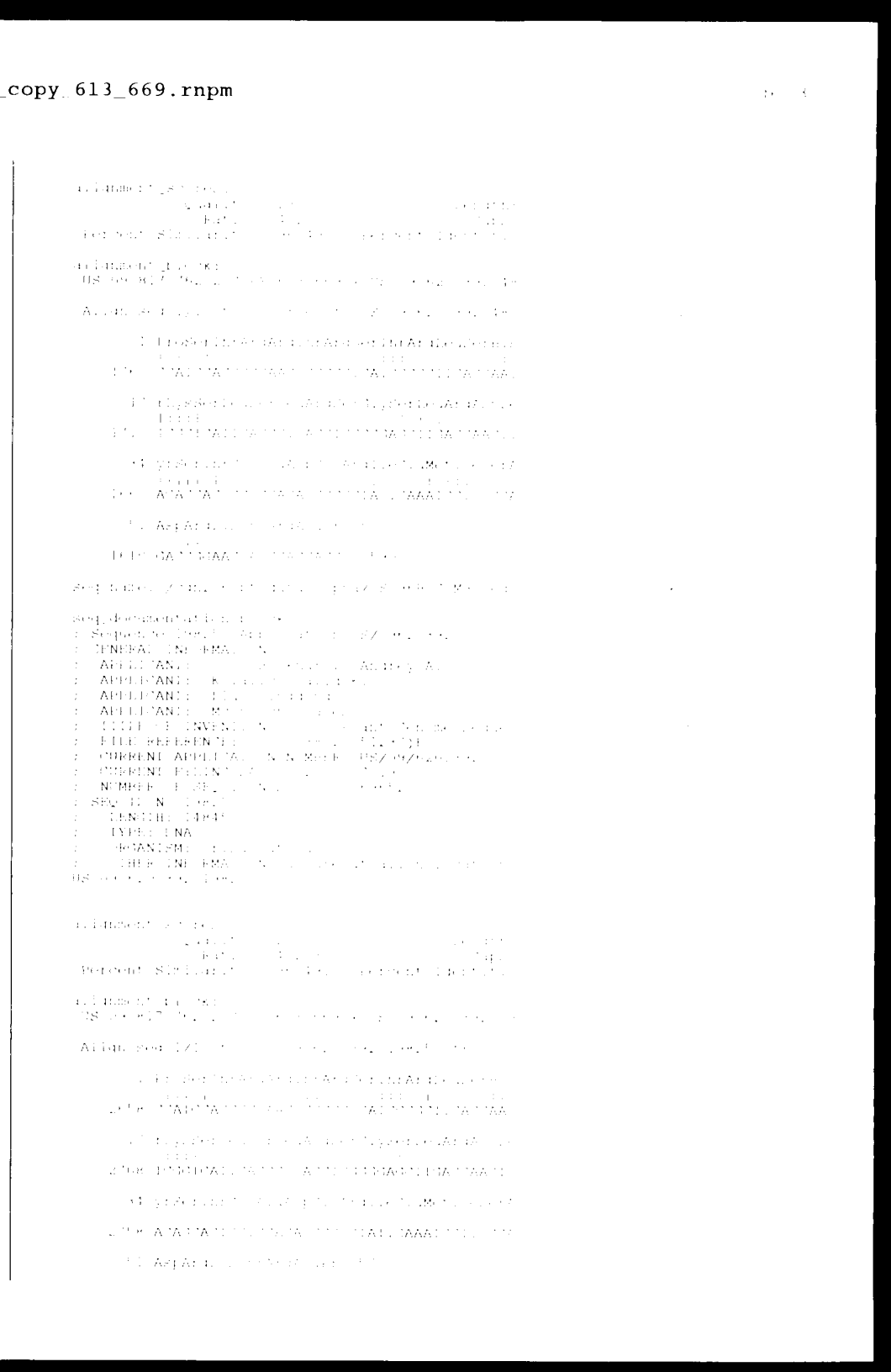
Align_Set 1/1 To reverse of: US-09-817-762-479 from: 1 to: 82430

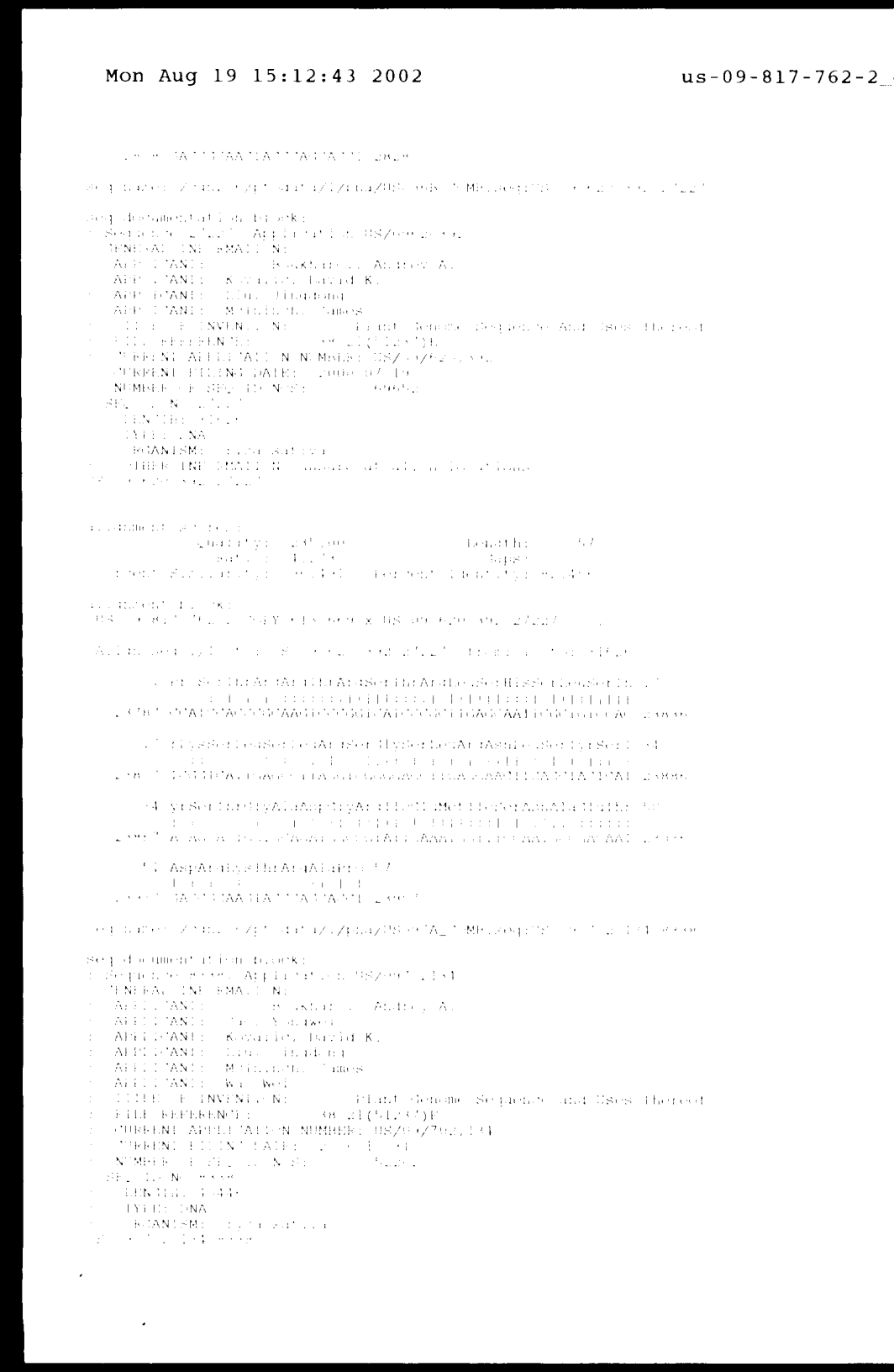
1 ProSerThrArgArgThrArgSerThrArgLeuSerHisSerLeuSerTh 17
||||||||||||||||||||||||||||||||||||||||||||||||||||
19667 cGGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGA 19716

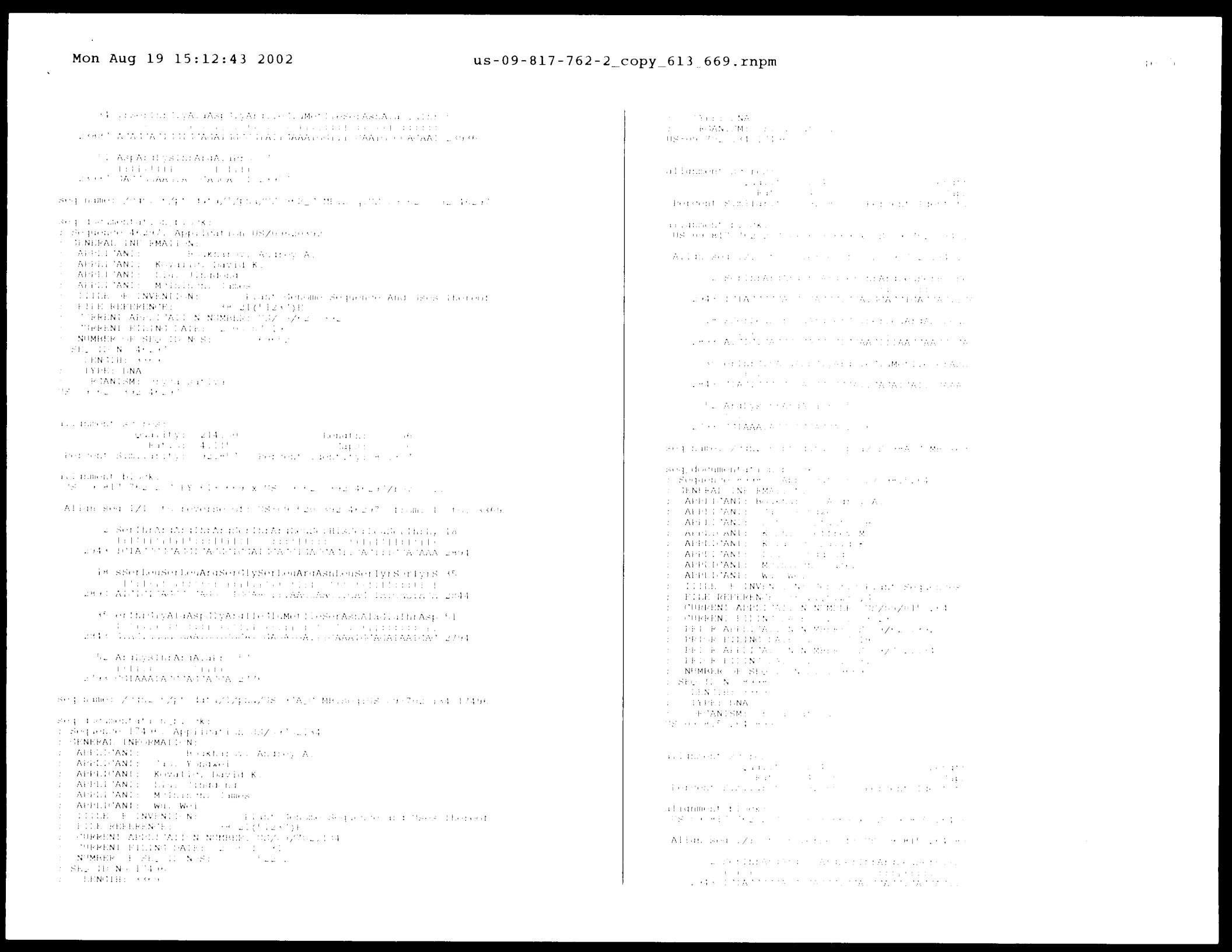
17 rLysSerLeuSerLeuArgSerGlySerLeuArgAsnLeuSerTyrSerT 34
||||||||||||||||||||||||||||||||||||||||||||||||||||
19717 cAAGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGAT 19766

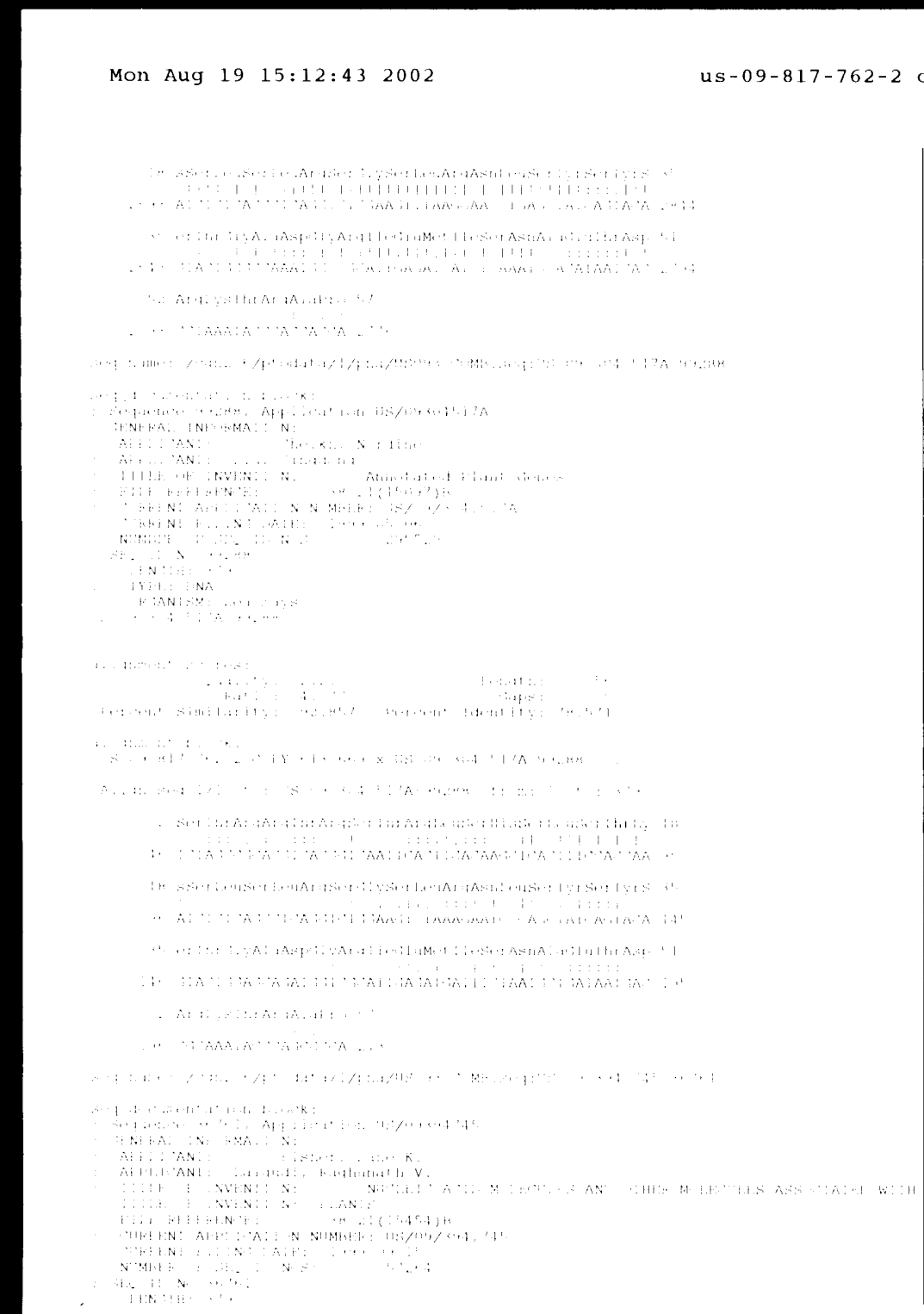
34 yrSerThrGlyAlaAspGlyArgProSerMetProSerAsnAlaGluThr 50
||||||||||||||||||||||||||||||||||||||||||||||||||
19767 ATGCCGAAAGACTCCTCGGCGCCTCAGCCTAAGCTCCTTAGCCTTCGAT 19816

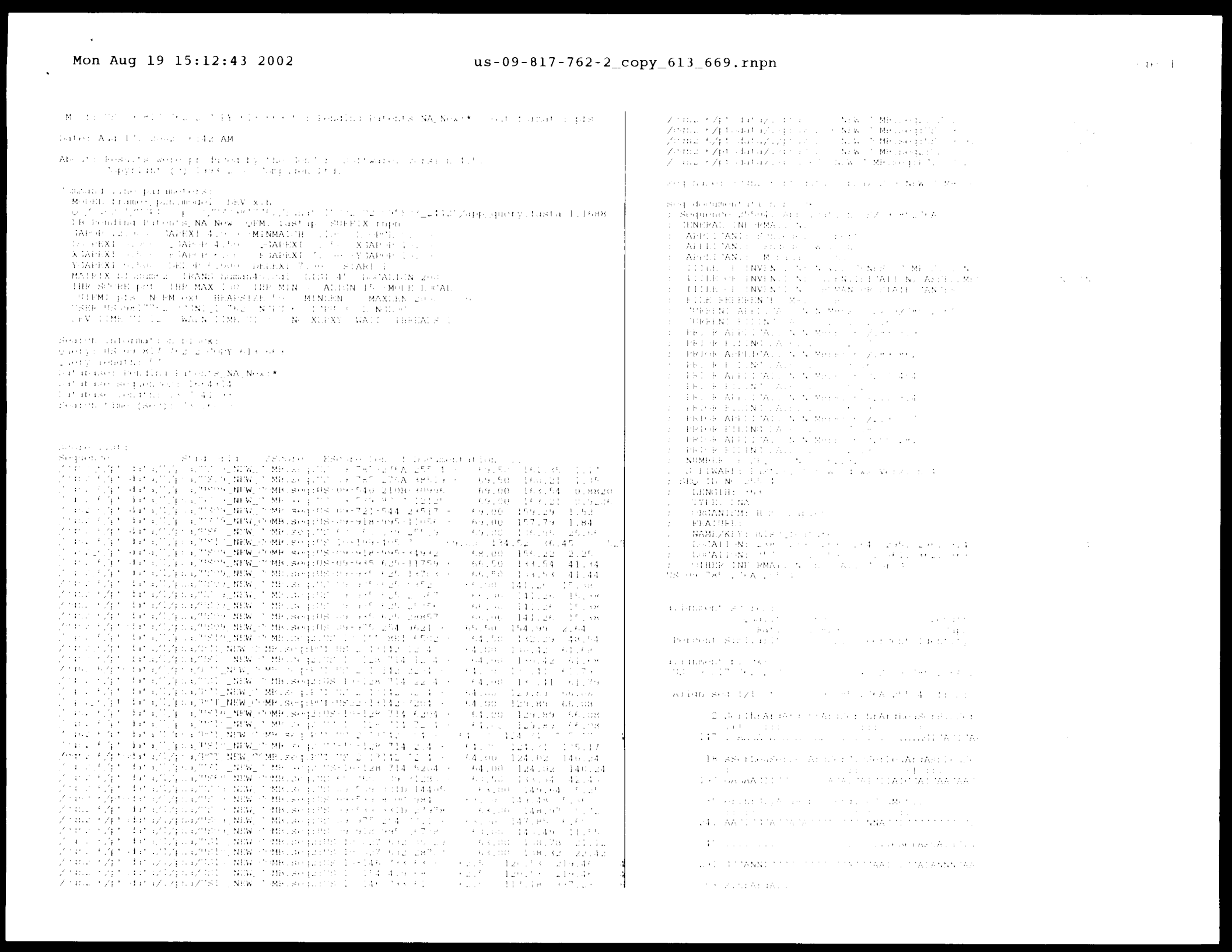


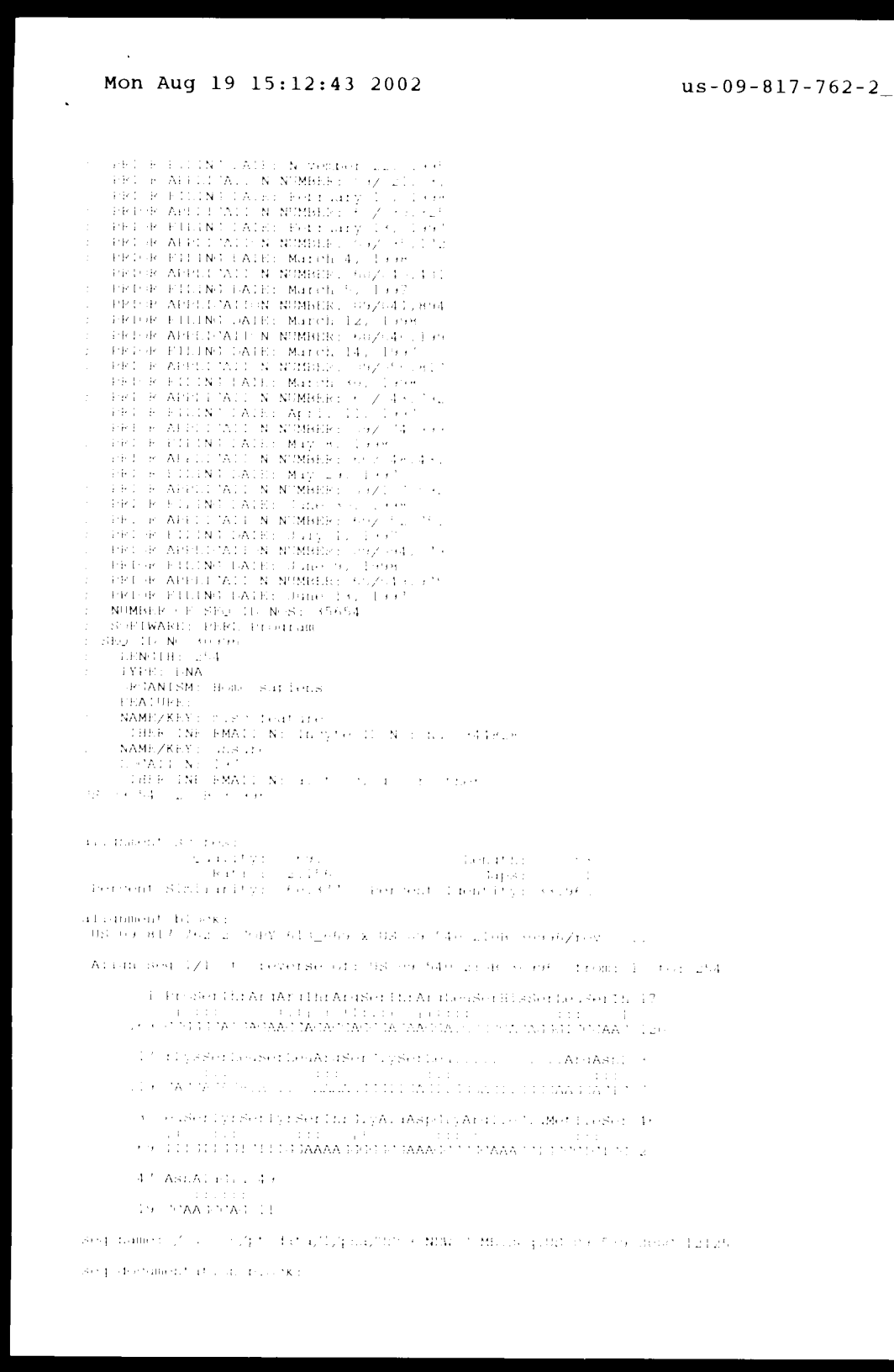


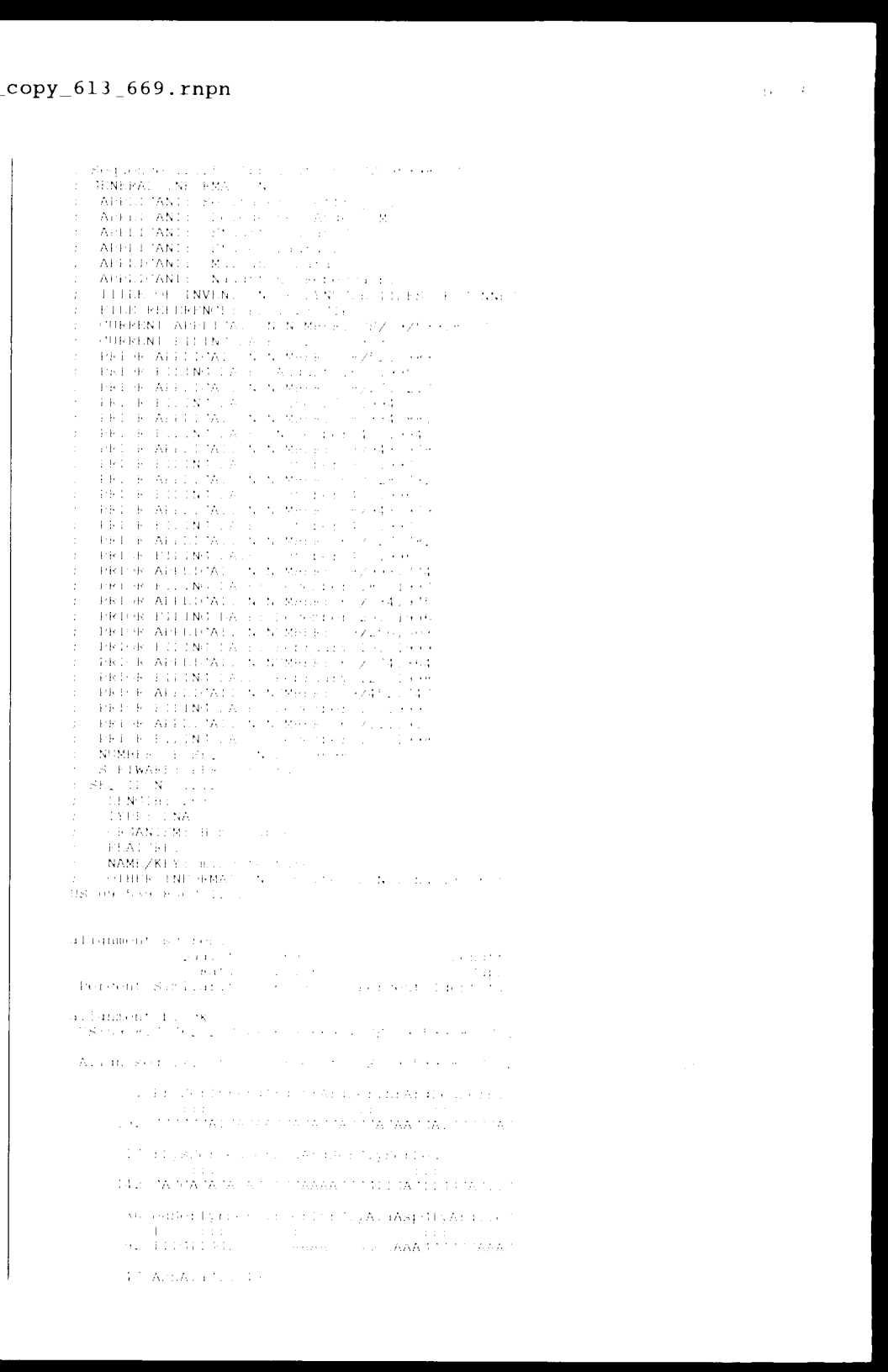












```
                  ||||||
      42 CAAGGCAG 49

Seq_name: Zoom_n/a_1_2/1/1_1:4/1/1_1:1/2000_NBA_1_DB_seq_08-09-721-544_23517

Seq_documentation_block:
; Sequence 23517, Application US/09721544
; GENERAL INFORMATION:
;   APPLICANT: Arterburn, Matthew
;   APPLICANT:   Nemerit, Yigal
;   APPLICANT:   Hamamadi, Ghazi
;   APPLICANT:   Hickey, Mark
;   APPLICANT:   Drake, Jim
;   APPLICANT:   Emmanuel, Rodaji
;   APPLICANT:   Eichmann, Carrie
;   APPLICANT:   Parikh, Brandy
;   APPLICANT:   Garcia, Veronica
;   APPLICANT:   Xu, Xiaolan
;   APPLICANT:   Hunter, Kelly
;   APPLICANT:   Bosse, Aaron
;   APPLICANT:   Benet, Leo
;   APPLICANT:   Kita, David
;   APPLICANT:   Cabot, Tom
;   APPLICANT:   Trova, Mimi
;   APPLICANT:   Tonelli, Michelle
;   APPLICANT:   Nguyen, Phuong
;   APPLICANT:   Neitz, Mavis
;   APPLICANT:   Pirotzky, Servando
;   APPLICANT:   Feist, Partha
;   APPLICANT:   Smith, Benjamin
;   APPLICANT:   Ikach, Jac
;   APPLICANT:   Tran, Lien
;   APPLICANT:   Verma, Ken
;   APPLICANT:   Yang, Pei
;   APPLICANT:   Yim, Kenneth
;   TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;   TITLE OF INVENTION:  From a cDNA Library of Fetal Liver-Spleen
;   FILE REFERENCE: Z2067IP
;   CURRENT APPLICATION NUMBER: US/09/721,544
;   CURRENT FILING DATE:  2000-11-21
;   PRIOR APPLICATION NUMBER: 09/015,129
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: 09/024,341
;   PRIOR FILING DATE: 1998-02-18
;   NUMBER OF SEQ ID NOS: 23517
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23517
;   LENGTH: 49
;   TYPE: DNA
;   ORGANISM: Homo sapiens
 Seq_id 7/21_544_23517

Alignment_scores:
             Quality:   99.00              Length:      53
               Ratio:   2.156               Gaps:       1
 Percent Similarity:   60.377   Percent Identity:  53.962

Alignment_block:
 1/1 09/817,762-2_COPY_613-669 x US-09-721-544-23517/rev  1

 Application 1/1  Reverse of: US-09-721-544-23517  From: 1  To: 496

        1 ProSerThrArgArgThrArgSerThrArgLeuSerSerSerLeuSerTh 17
          ||||||  |||||||||||||||||||||  |||  |||  ||
      496 CCGGGCGTCTGCGGCTCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCC 447

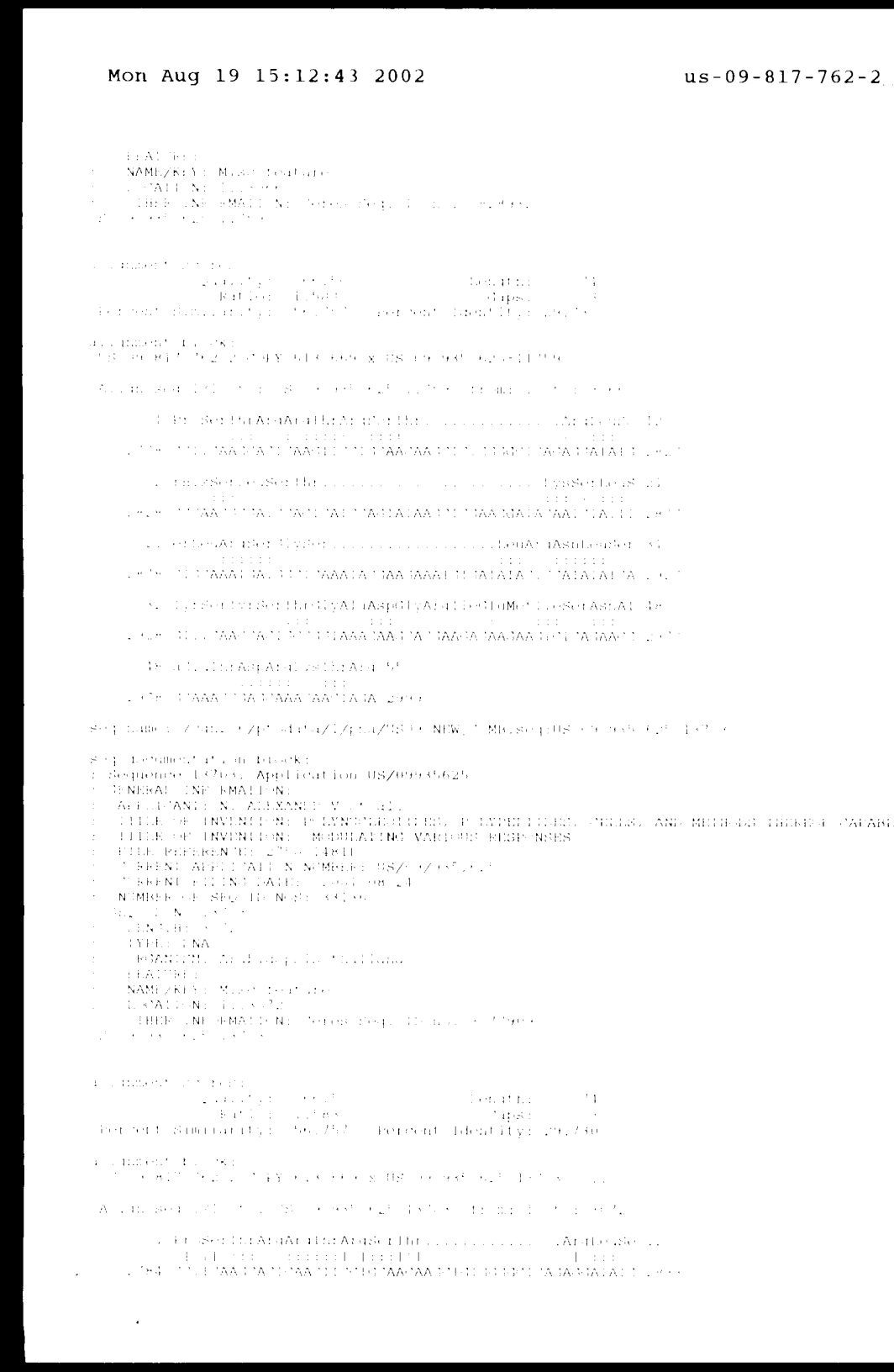
       17 rlysSerLeuSerLeuArgSerGlySerLeu.....ArgAsnL 49
          |  |||  |||  |||  |||  |||  |||           |||||
      446 CGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGC 397

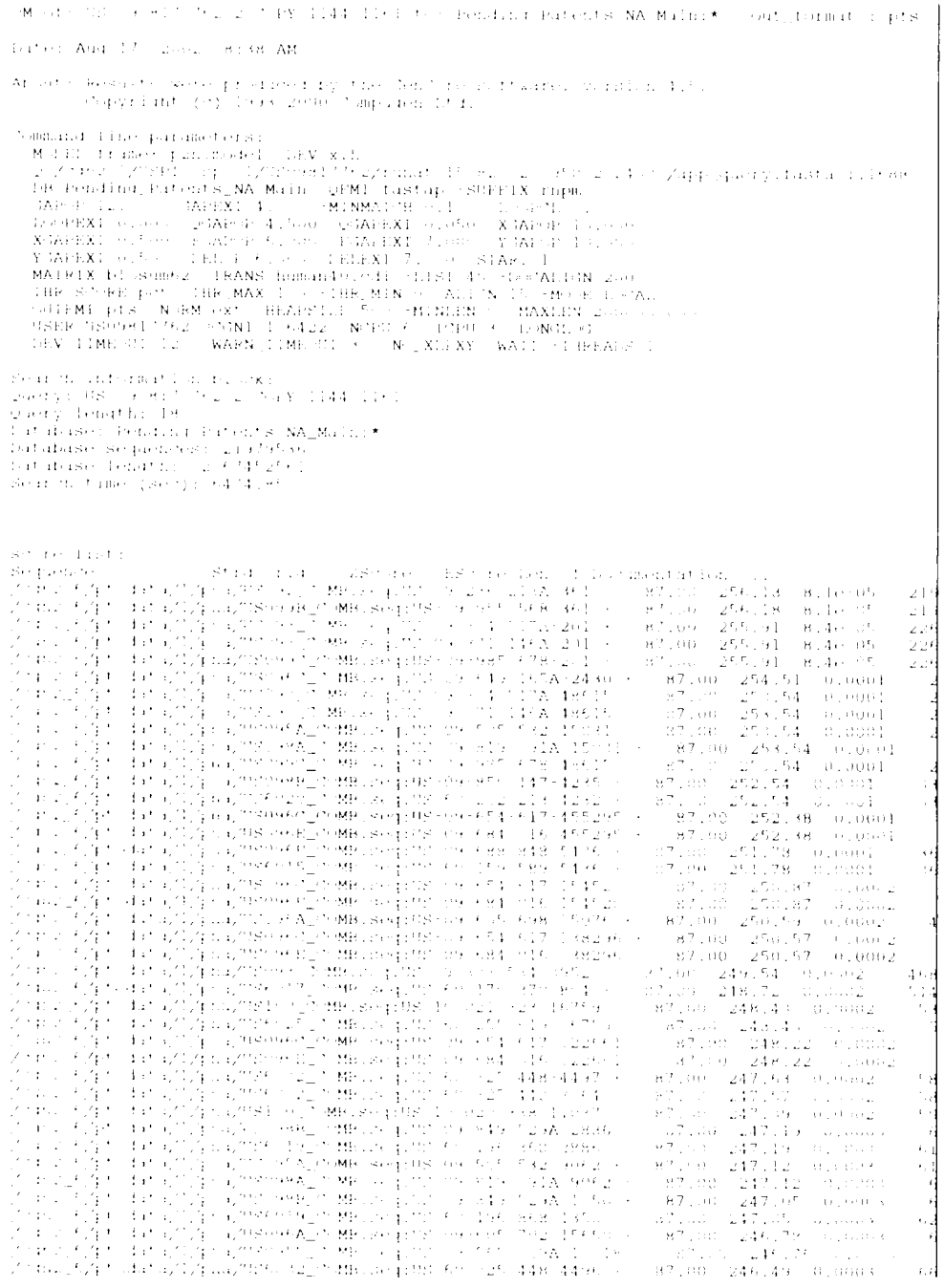
       50 euSerTyrSerTyrSerThrGlyAlaAspGlyArgIleGlnMetIleSer 46
          ||  |||||  |||  ||||||  |||||           |||||
      396 CTCTCCTACTCCTACTCCACCGGCGCCGACGGCCGCATCCAGATGATCTCC 347

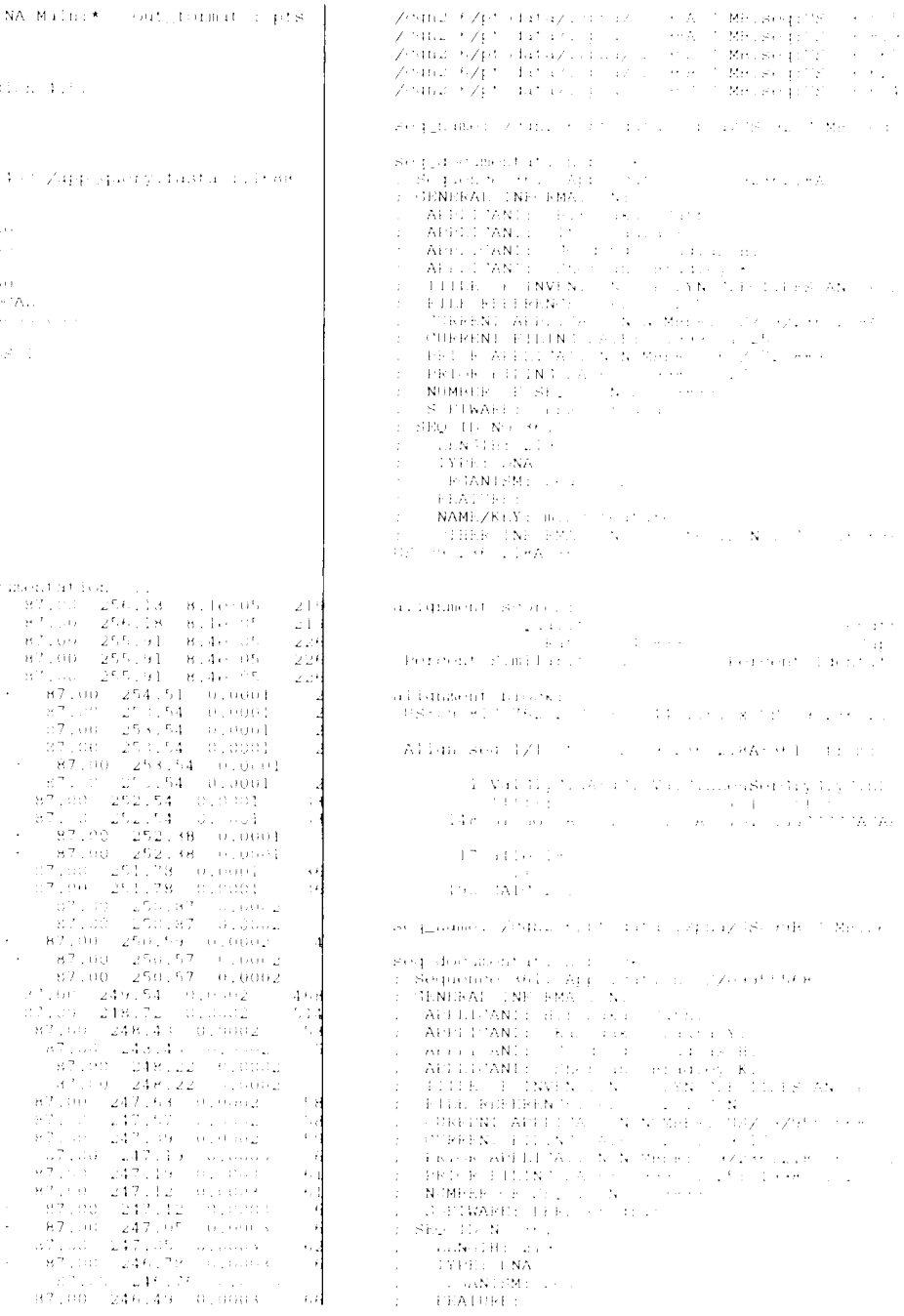
       47 AsnAlaGlu 49
          ||||||
      82 CCAAGCCAG 74

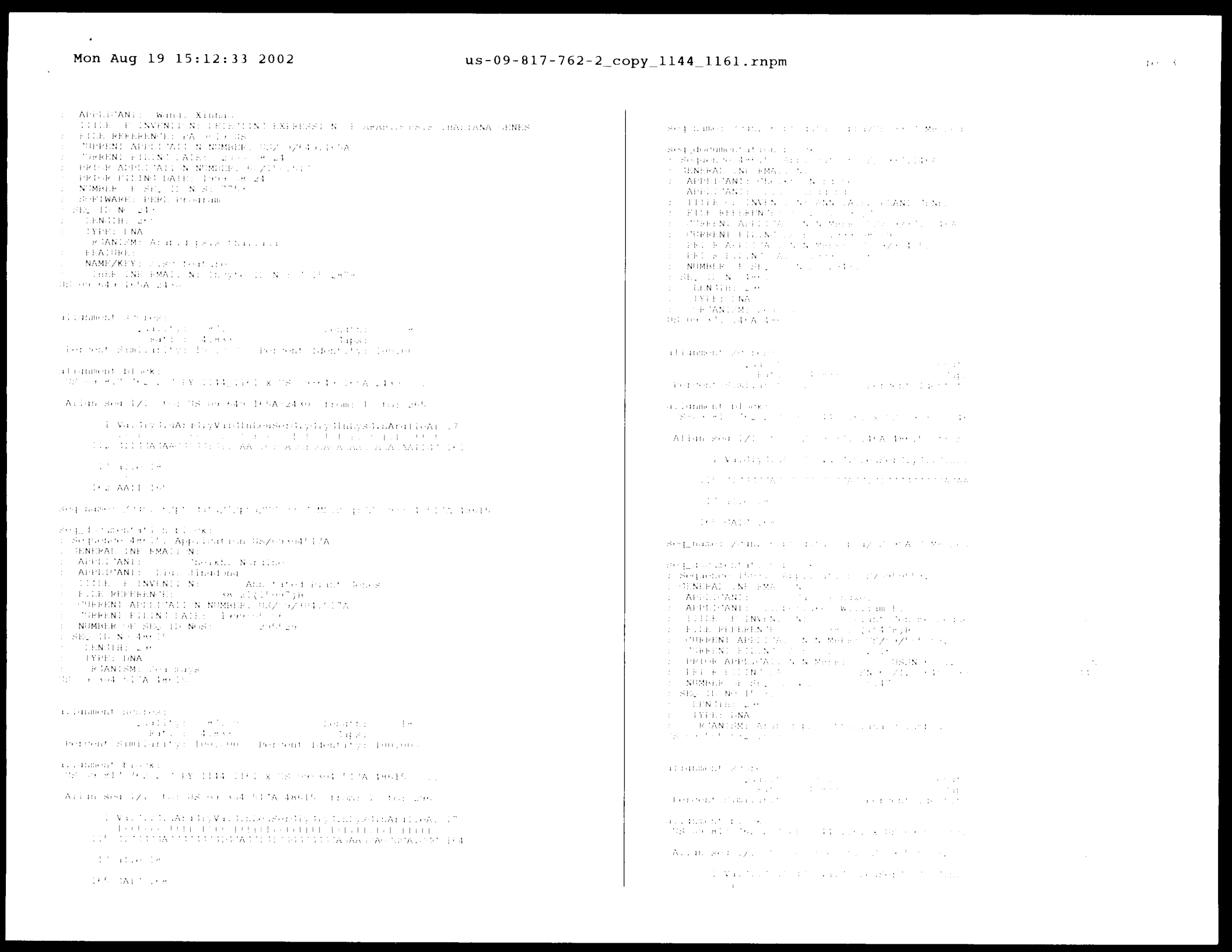
Seq_name: Zoom_n/a_1_2/1/1_1:4/1/1_1:1/2000_NBA_1_DB_seq_08-09-721-544_23517

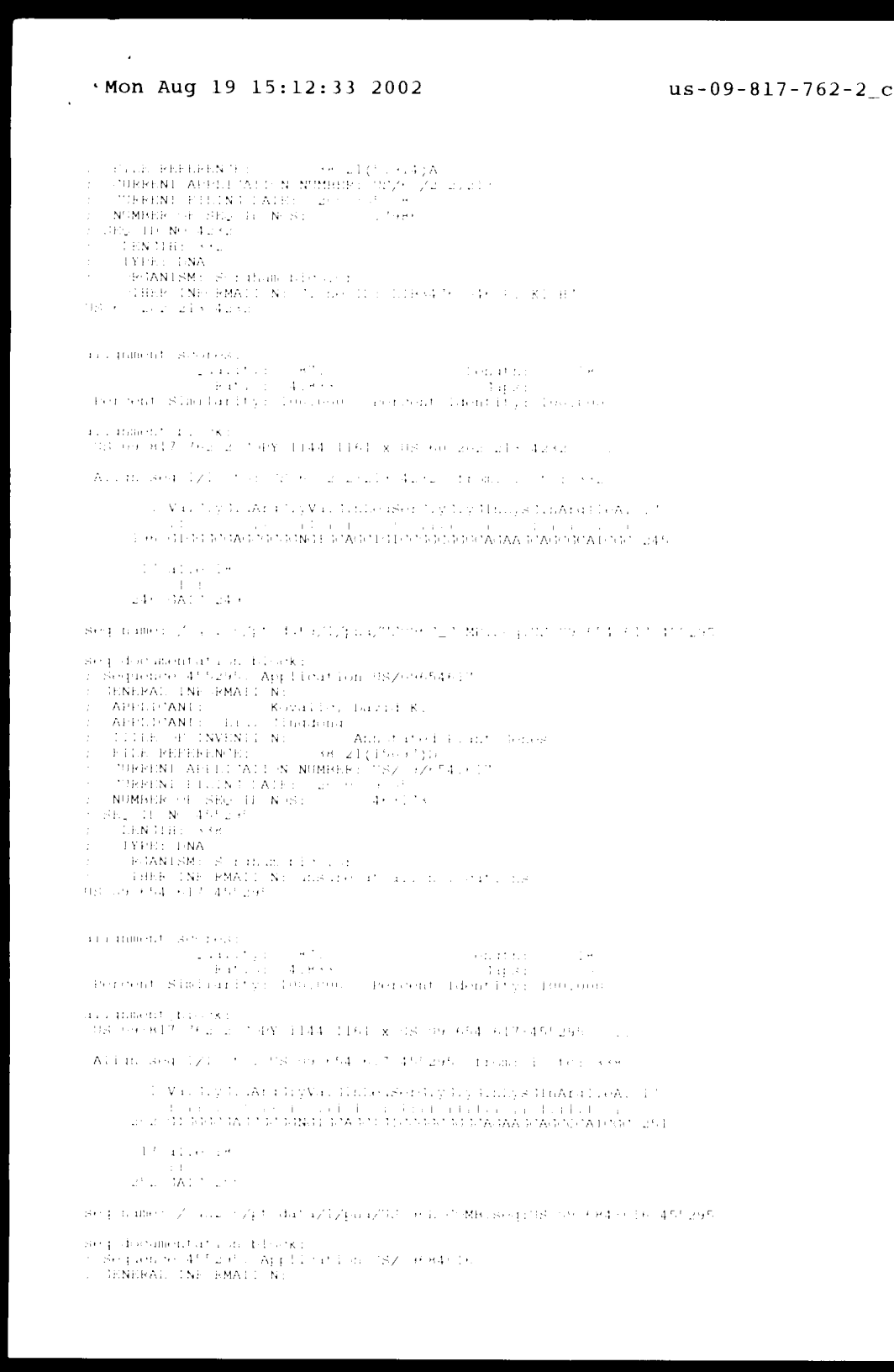
Seq_documentation_block:
; Sequence 23517, Application US/09721544
; GENERAL INFORMATION:
;   APPLICANT: Cao, Yongwei
;   APPLICANT:  Chen, Xiaofeng
;   APPLICANT:  Goldman, Barry S.
;   APPLICANT:  Hinkle, Gregory J.
;   APPLICANT:  Slater, Steven C.
;   TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;   TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;   FILE REFERENCE: 38.10(5202)A
;   CURRENT APPLICATION NUMBER: US/09/721,544
;   CURRENT FILING DATE:  2002-02-21
```

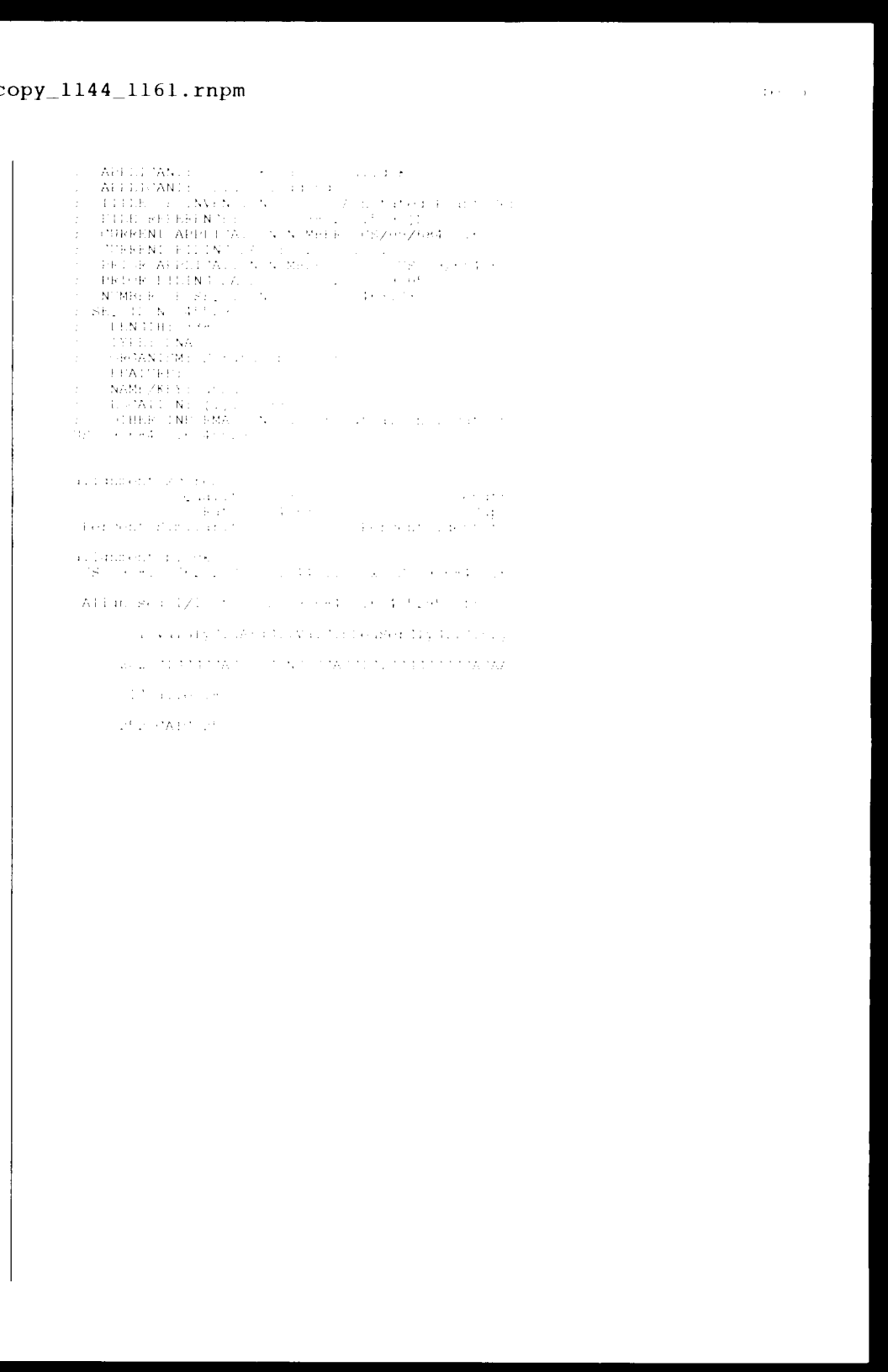


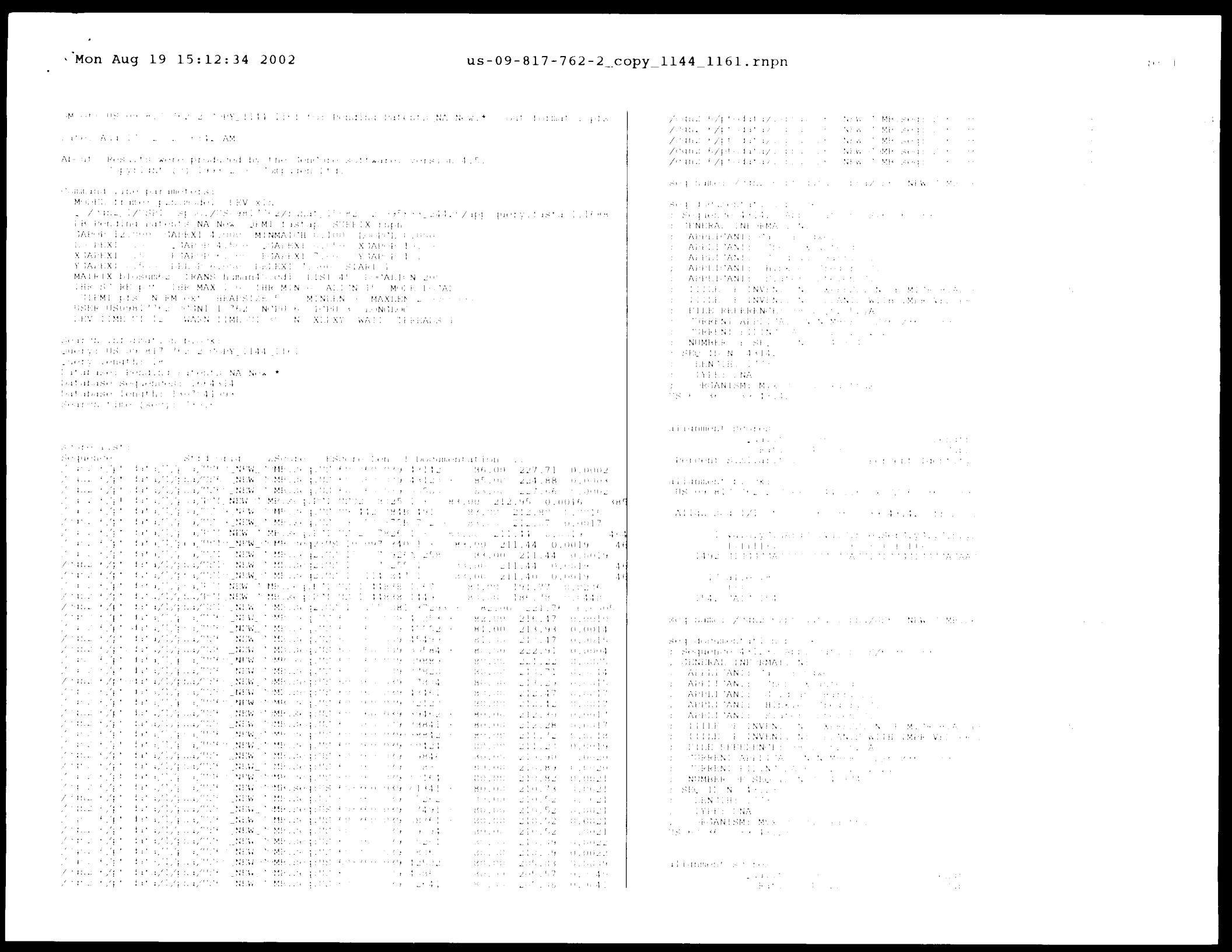


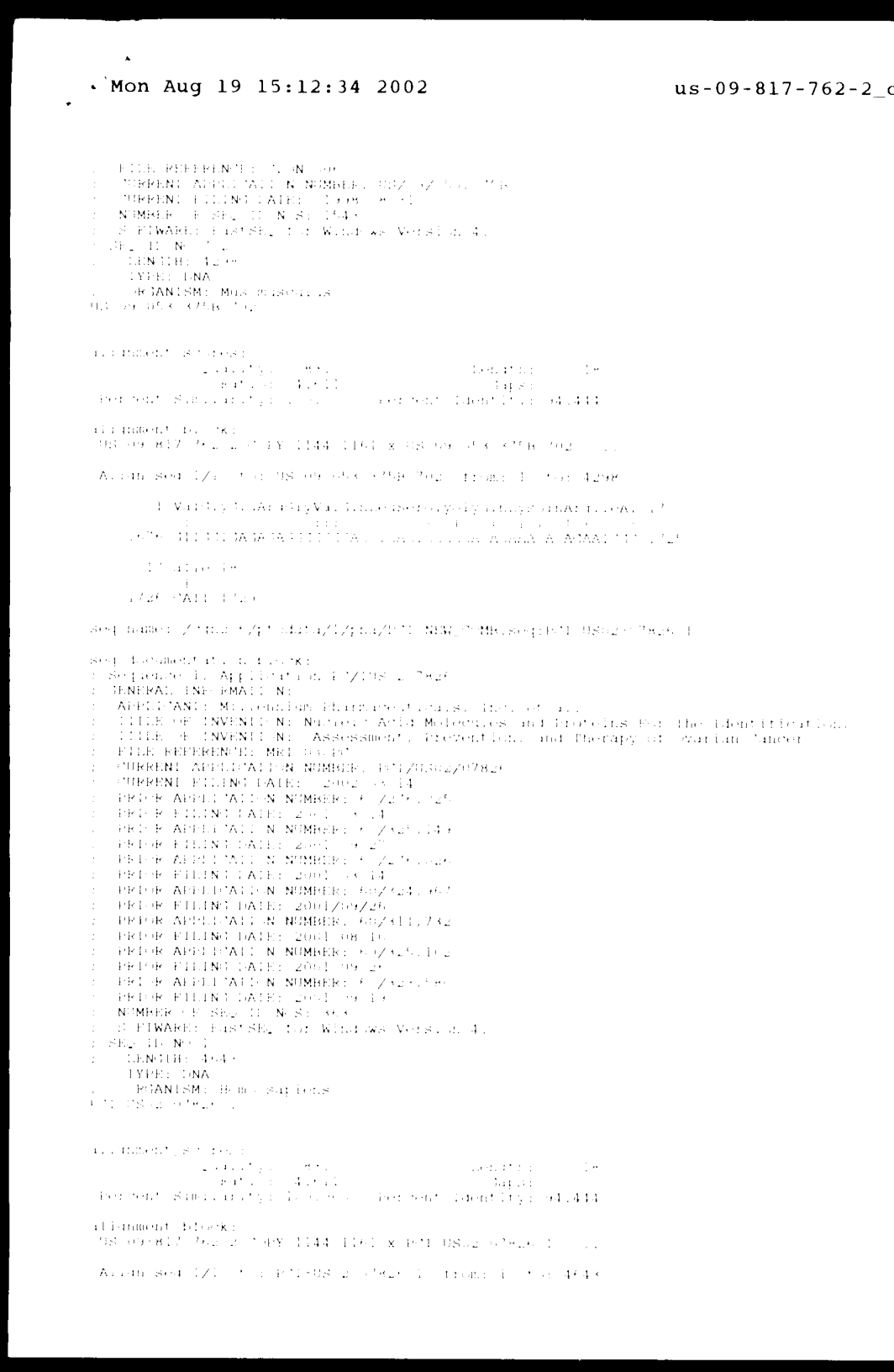


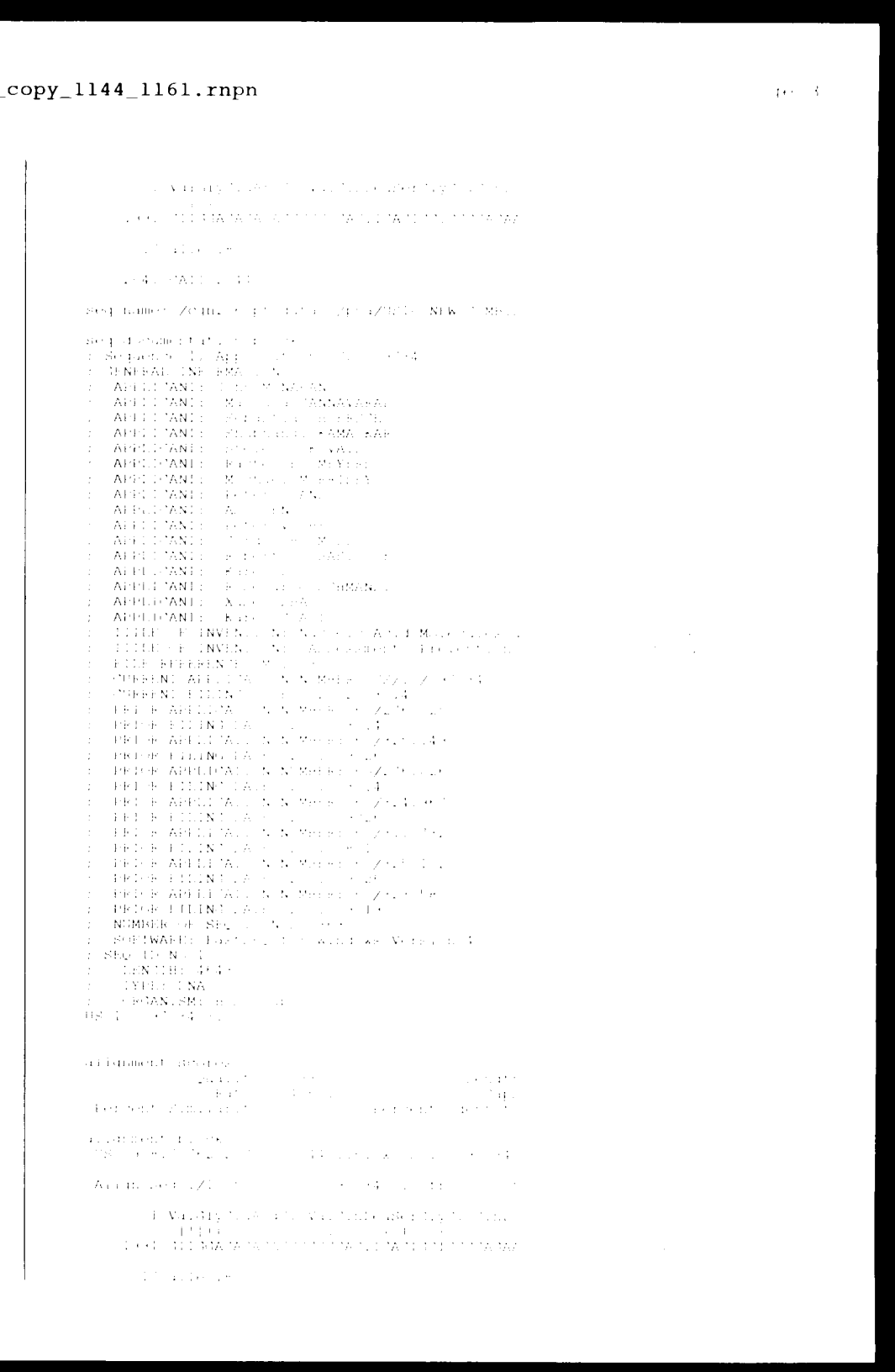


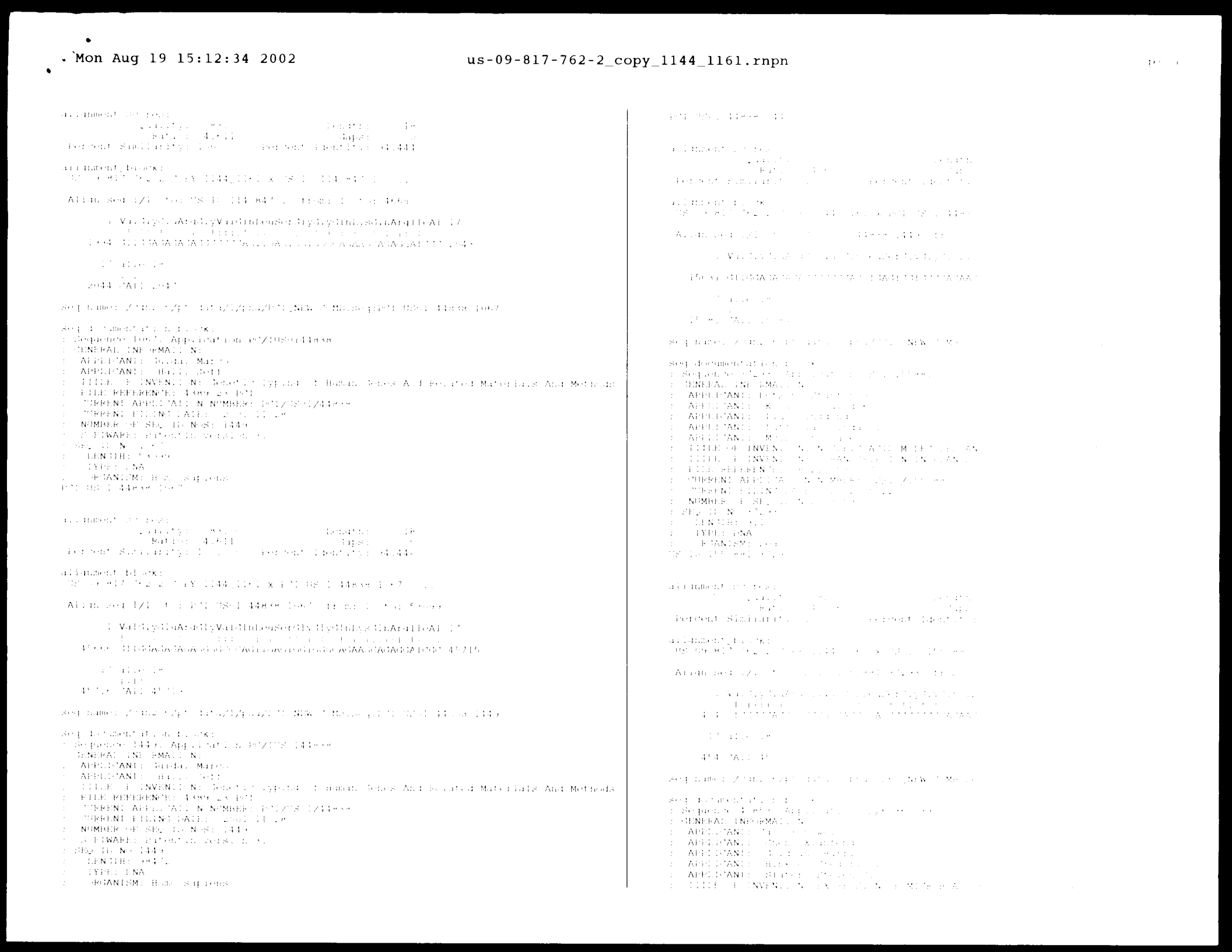


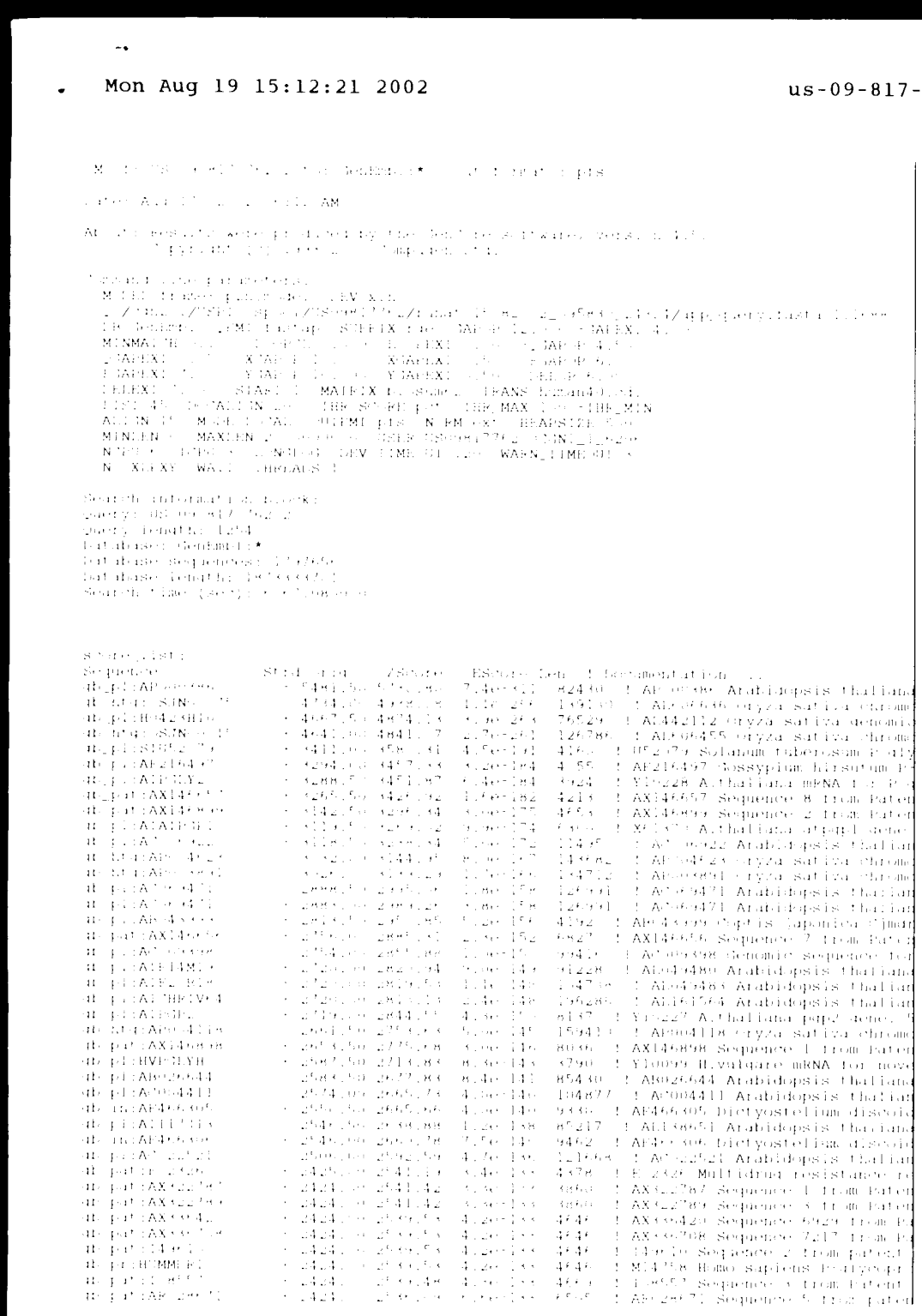












..22306,23182..23236,23418..23572))
/note="gene_id:MLD15.2"
/codon_start=1
/evidence=not_experimental
/product="P-glycoprotein; multi-drug resistance related; ABC transporter like protein"
/protein_id="BAB02129.1"
/db_xref="GI:9294227"
/translation="[illegible] HRLSTIRGVDCIGVIQDGRIVEQGSHSELVSRPEGAYSRLLQLQTHRI"

CDS complement(29421..35755)
/note="emb|CAB75612.1
gene_id:MLD15.3
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental

CDS join([illegible])
/note="gb|AAD[illegible]
gene_id:MLD15.4
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02130.1"
/db_xref="GI:9294228"
/translation="[illegible] LSNGKIS[illegible]VGGTKKN"

CDS join([illegible],39942,39761..40129,40210..40340,40433..40541, 40603..40766,41015..41669,41753..42054,42148..42181)
/note="gene_id:MLD15.5
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02131.1"
/db_xref="GI:9294229"
/translation="[illegible]"

CDS complement(42992..45127)
/note="gene_id:MLD15.6"
/codon_start=1
/evidence=not_experimental
/product="disease resistance protein like"
/protein_id="BAB02132.1"
/db_xref="GI:9294230"
/translation="[illegible] [illegible]HWCIK[illegible]TIH"

CDS complement(join(49156..49243,49323..49432,50436..[illegible], 50623..50701))
/note="gene_id:MLD15.7"
/codon_start=1
/evidence=not_experimental
/product="60S ribosomal protein L34"
/protein_id="BAB02133.1"
/db_xref="GI:9294231"
/translation="MV[illegible] VKKVEKIQKAKKKIAPKS"

CDS join(57725..57861,58188..58317,58434..59142)
/note="gene_id:MLD15.8"
/codon_start=1
/evidence=not_experimental
/product="MYB family transcription factor like protein"
/protein_id="BAB02134.1"
/db_xref="GI:9294232"
/translation="MVRP[illegible] [illegible]ILLDENINFF"

CDS complement(63869..69193)
/note="gene_id:MLD15.9"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="non-LTR retroelement reverse transcriptase like protein"

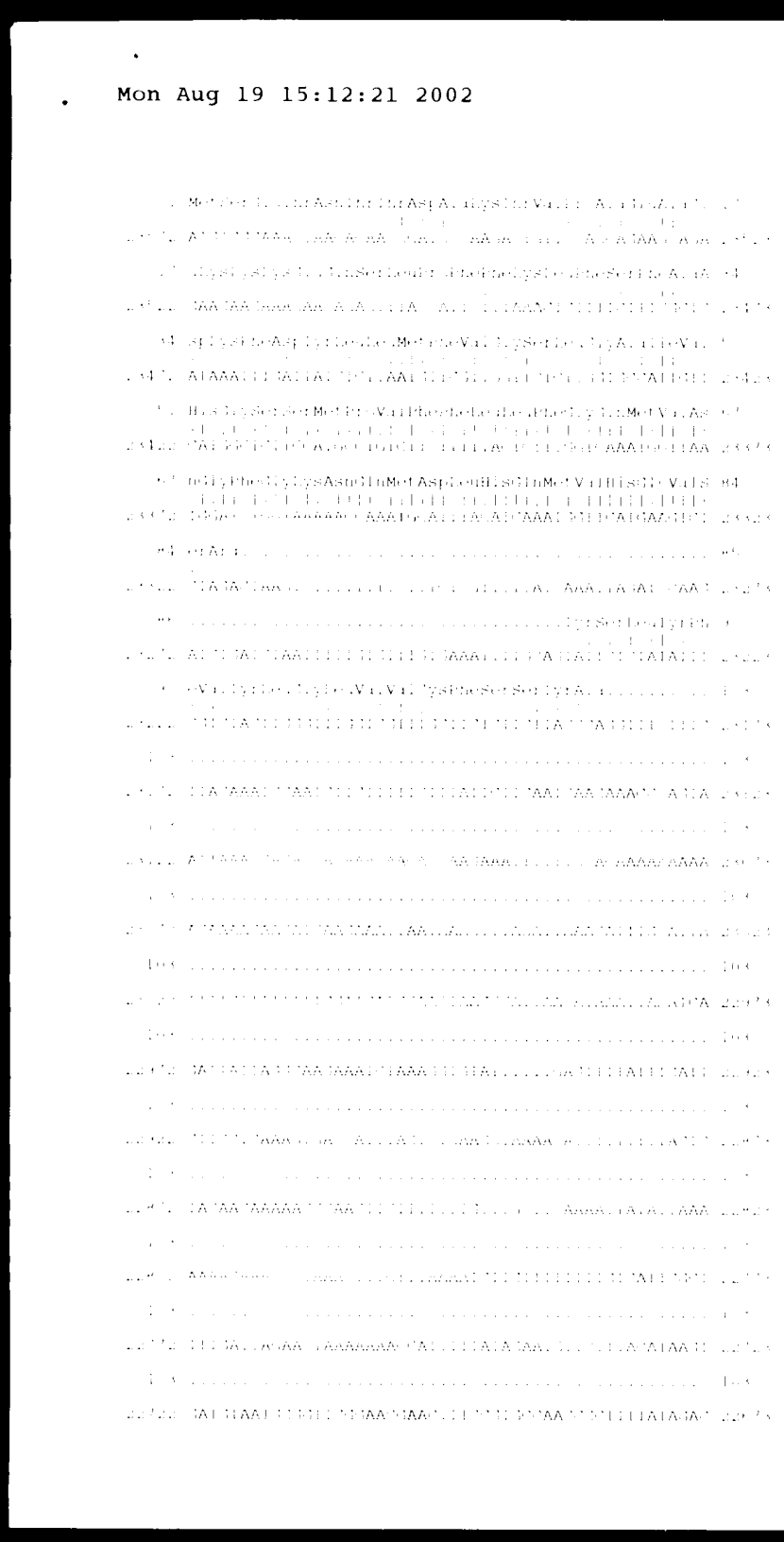
CDS 71380..71662
/note="gb|AAF15936.1
gene_id:MLD15.10
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02135.1"
/db_xref="GI:9294233"
/translation="MLKL[illegible]"

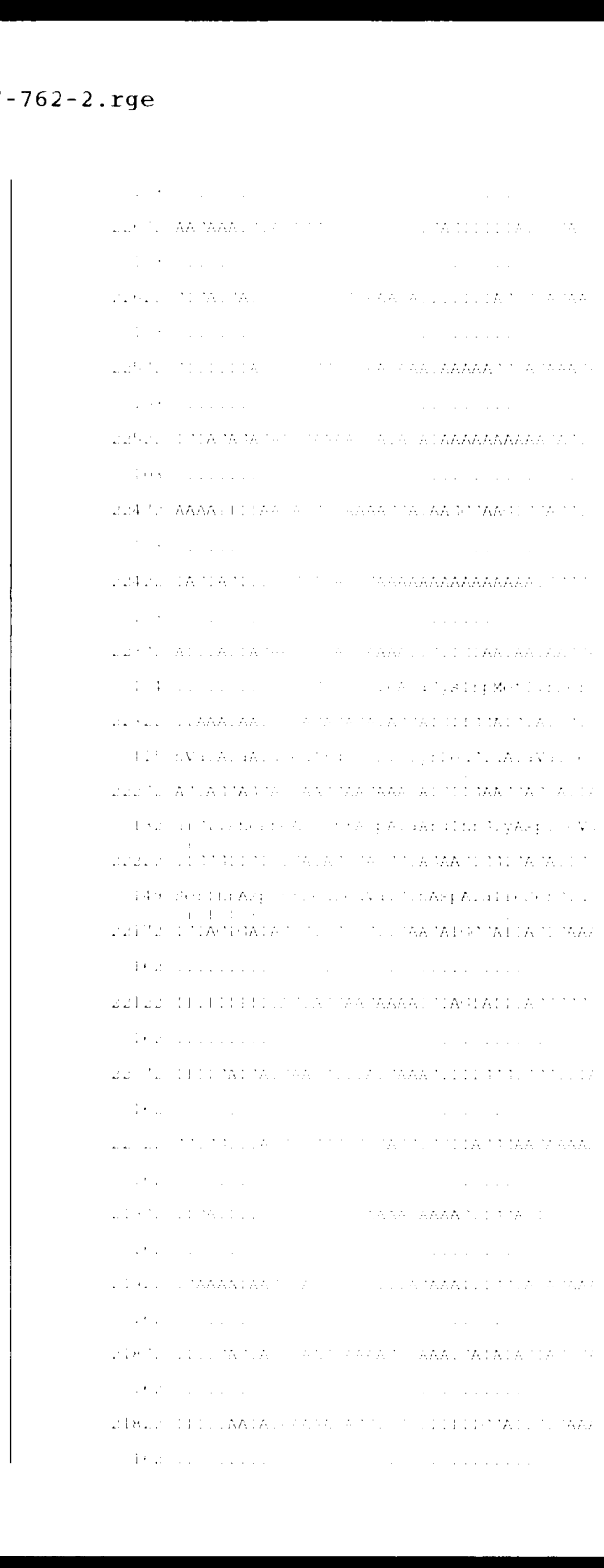
misc_feature complement(74370..75514)
/note="contains similarity to retrotransposon reverse transcriptase"

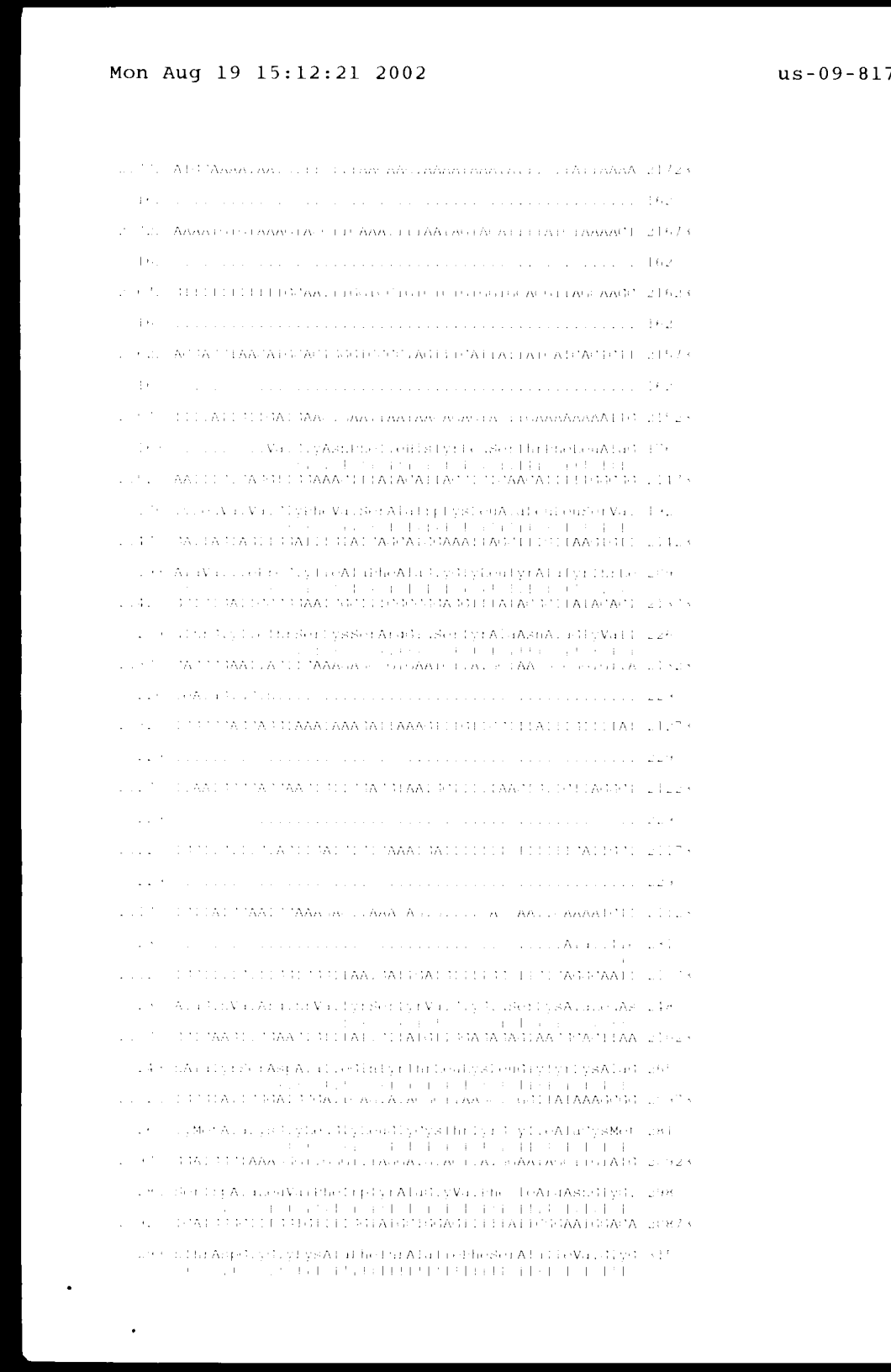
alignment_scores:
Quality: 5488.50 Length: 2338
Ratio: 4.391 Gaps: 17
Percent Similarity: [illegible] Percent Identity: [illegible]

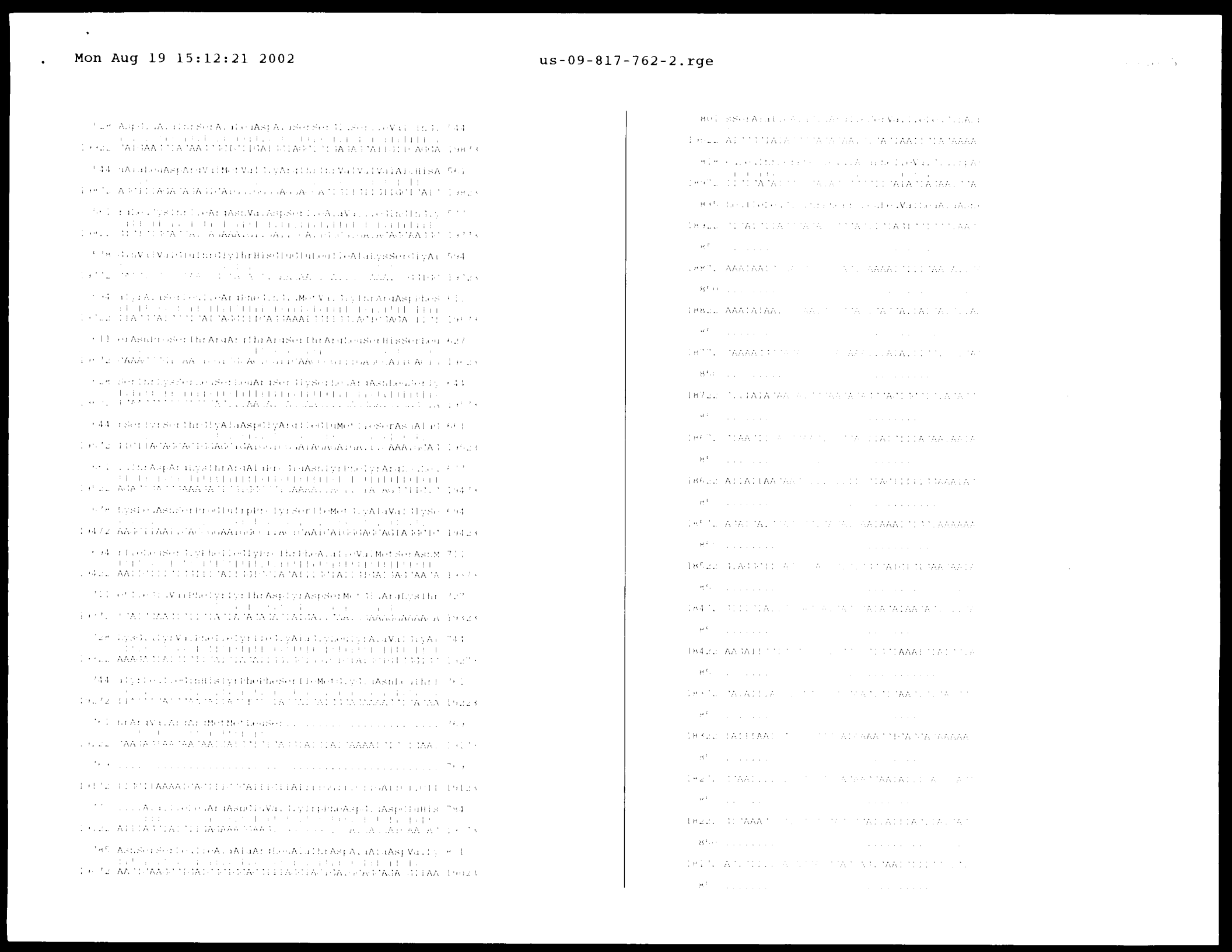
alignment_block:
US-09-817-762-2 x AP000386 ..

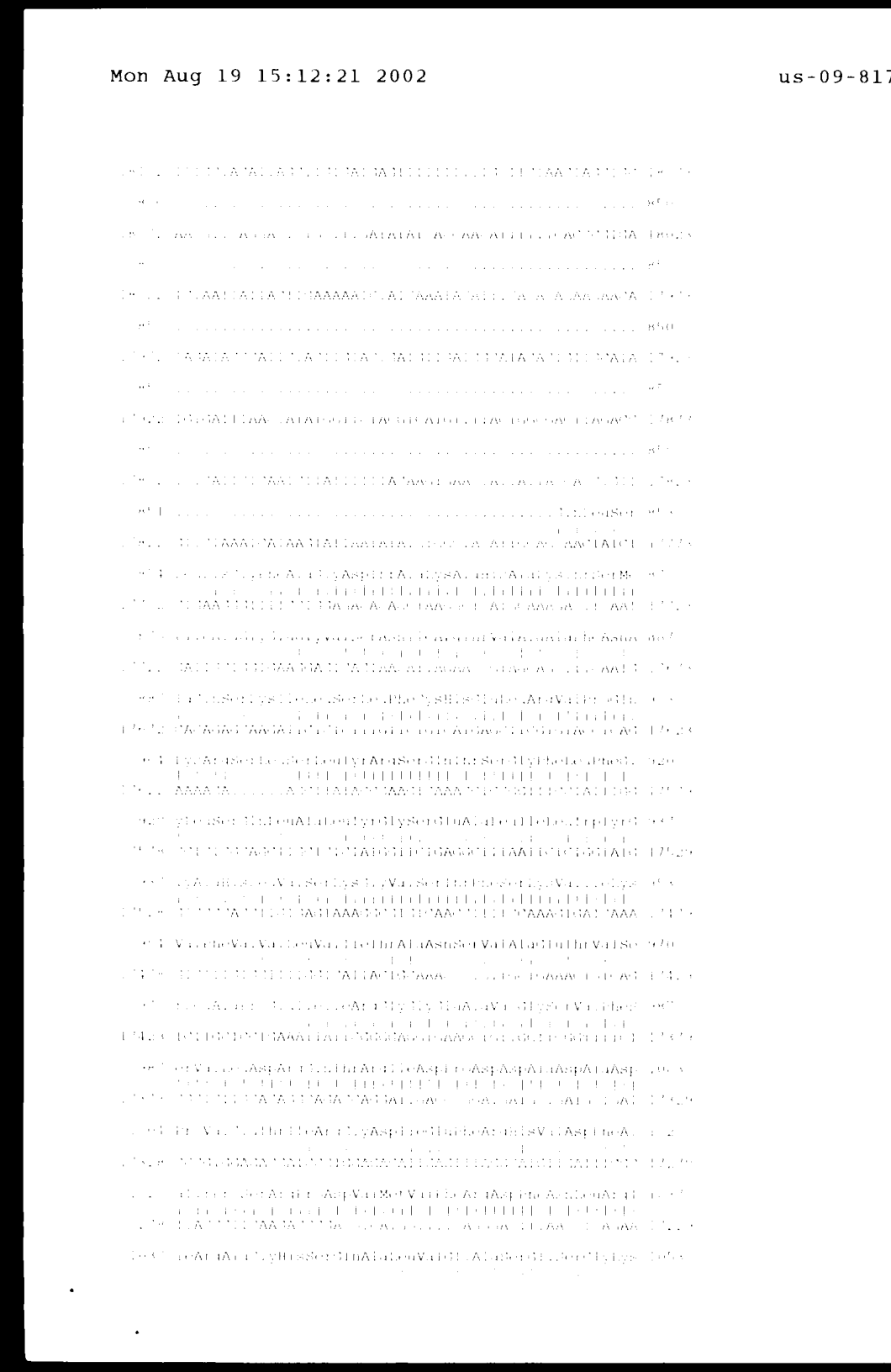
Align seq 1/1 to reverse of: AP000386 from: 1 to: 82430

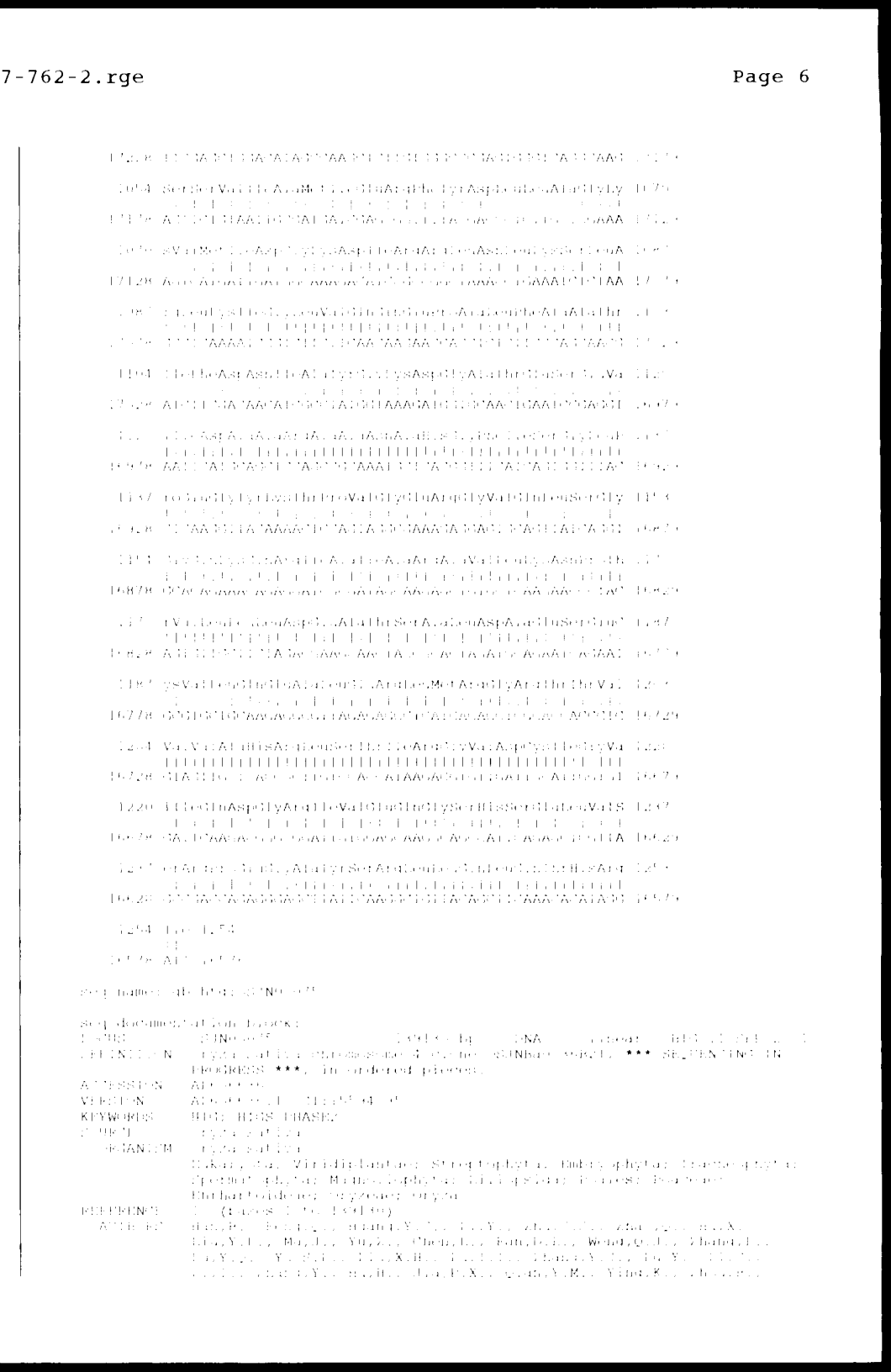












Zhou,J.H., Bao,F., Zhang,L., Wu,M., Zhang,F.Q., Guan,J.P., Fu,G.,
Wang,S.Y., Kong,S.X., Li,J., Liu,W., Li,W.Q., Zhu,G.F., Du,Y.F.,
Fu,Y., Yin,H.F., Zhang,Y., Cai,Z., Sun,J., Kang,L., Chen,X.Y.,
Shi,J.Y., Shi,Y., Bu,J.F., Zhang,X.L., Zhang,W., Wang,L.,
Cui,Y.W., Shen,H.B., Li,J., Zhou,S.F., Ni,J., Zhu,F.H. and
Hou,L.F.

TITLE Direct Submiss.n

JOURNAL Submitted (28-JUL-2) Bob Bin, National Center for Gene Research,
Chinese Academy of Sciences, 500# Cai Lun Road, Shanghai 200233,
CHINA, E-mail enquiries: bhan@ncur.ac.cn, Clone requests:
bhan@ncur.ac.cn.

REMARK Physical mapping (high-density) of human DNA: chromosome 4, BAC
7 (set: SCNBa 9B2).

COMMENT Web site: http://www.ncur.ac.cn
-- Summary Statistics
Assembly performed by
* NOTE: This is a "working draft" sequence
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

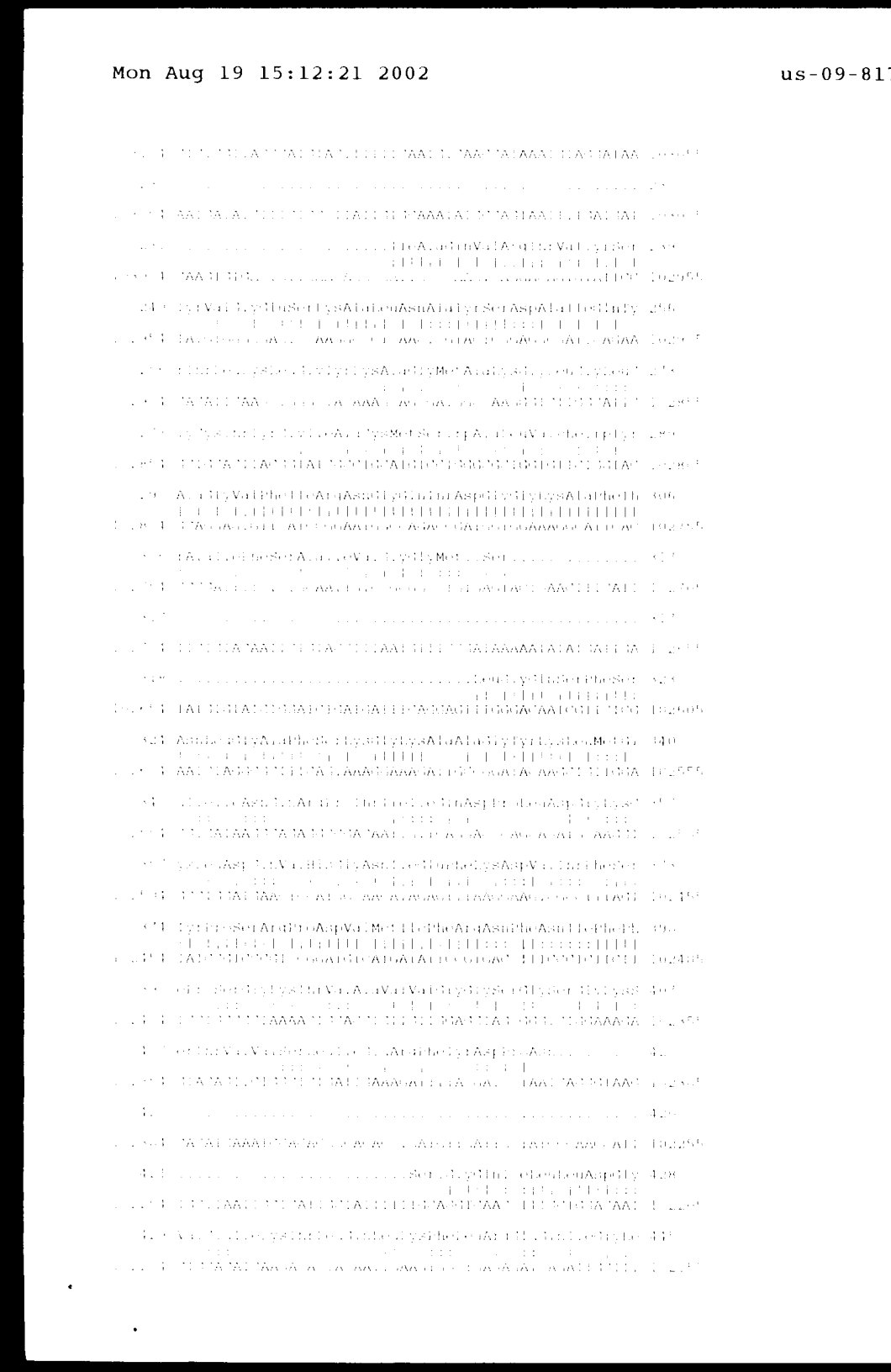
FEATURES Location/qualifiers
source 1..150150
/chromosome "4"
/clone "SCNBa 9B2"
/clone_lib "PCI SCNBa"
BASE COUNT 39111 a 31944 c 32 g 39113 t 950 others
ORIGIN

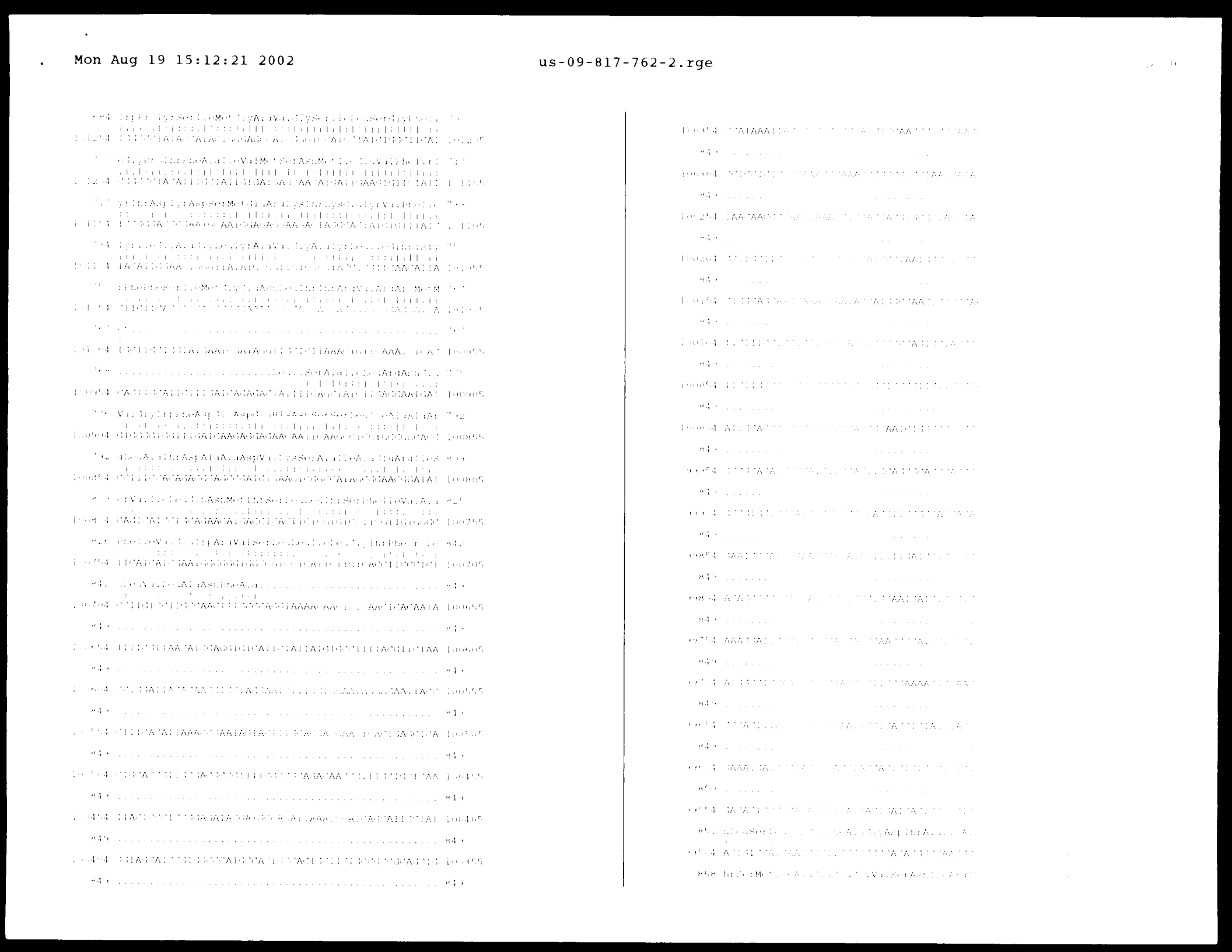
Alignment Block:
CS 09-817 762-2 2 X SCNBa 9B2/rc

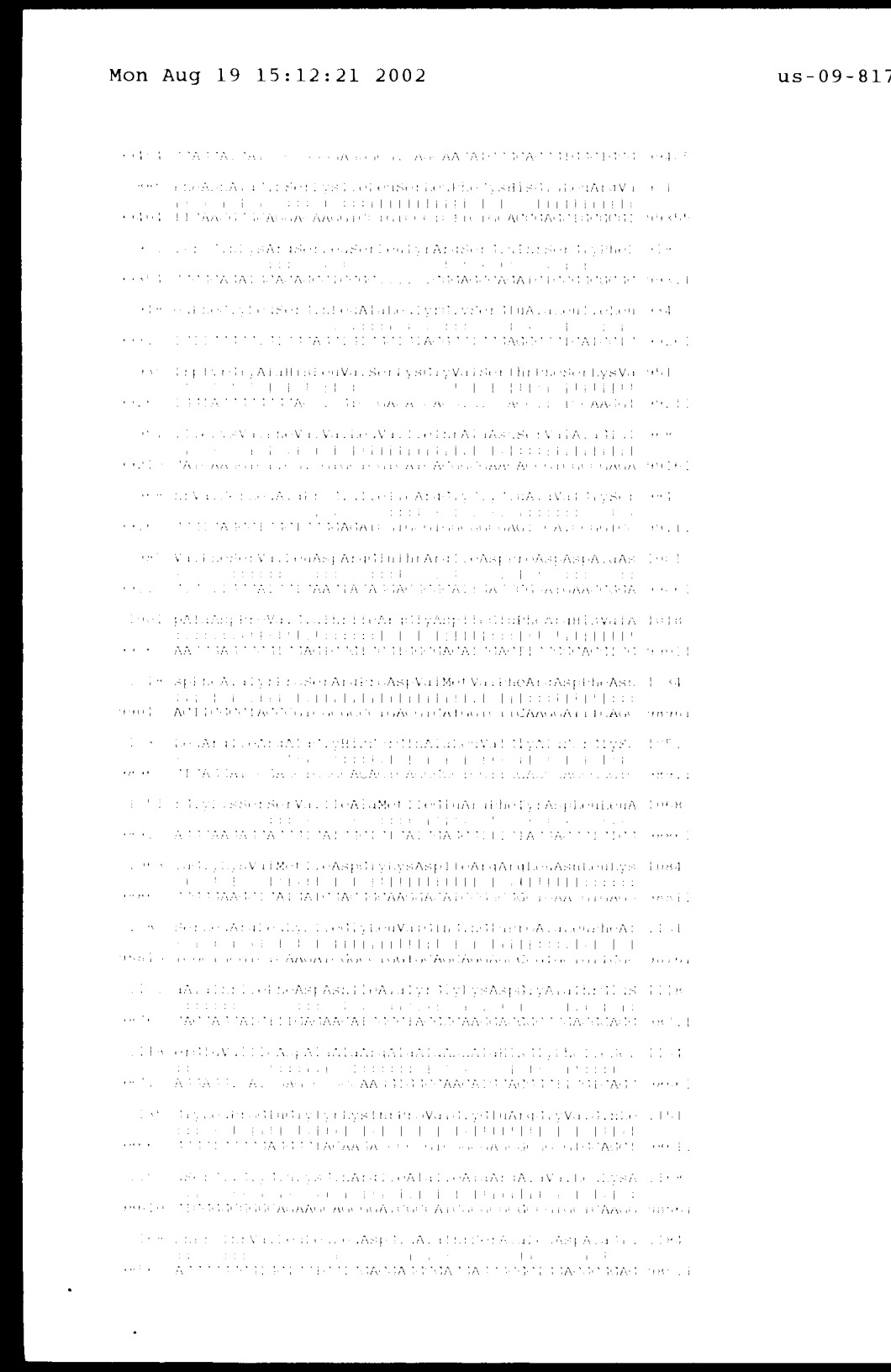
Alignment of: CS 09-817 762-2 2 X SCNBa 9B2/rc to all others
Minimum Match 10%
Shading mode Similarity
Percent Identity 51.4%

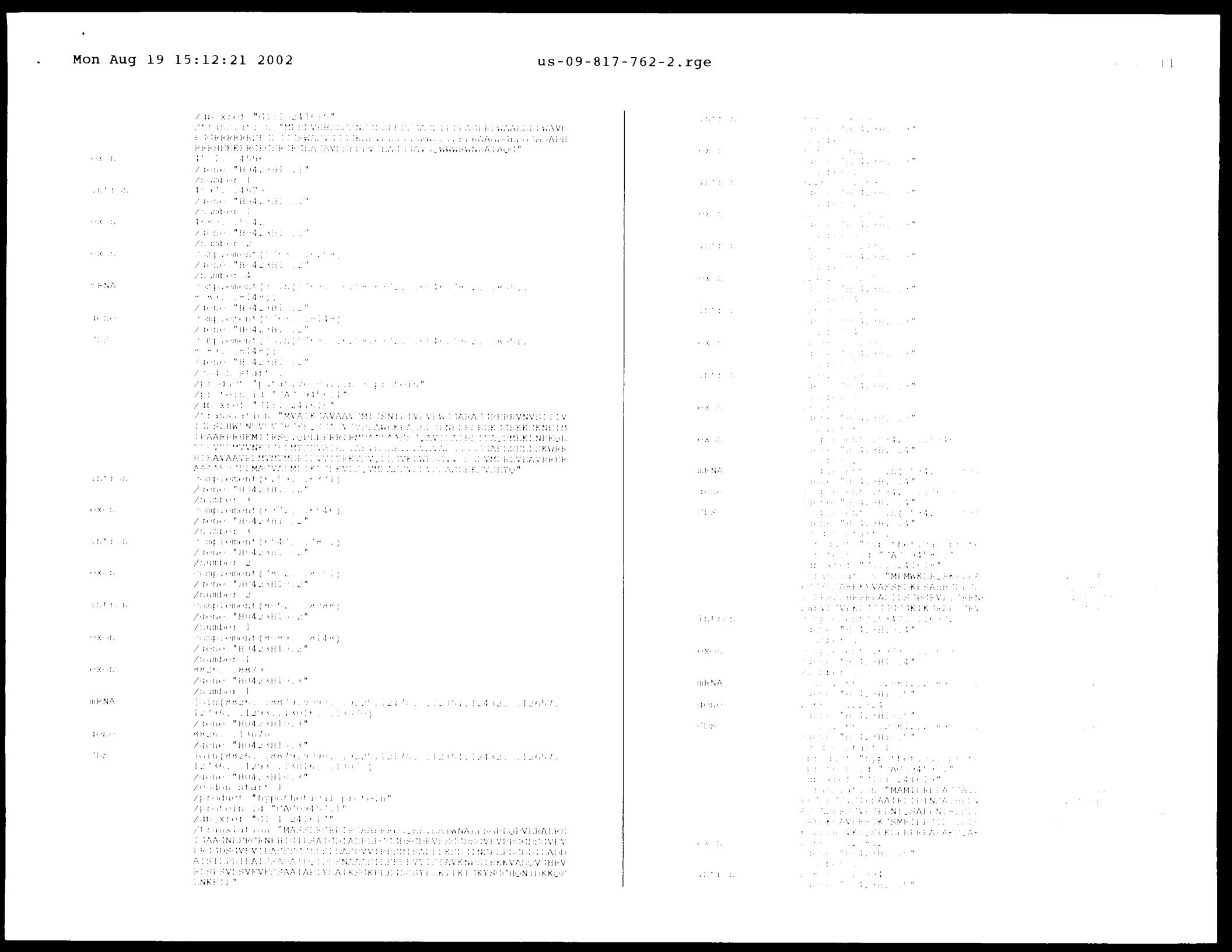
Alignment Block:
CS 09-817 762-2 2 X SCNBa 9B2/rc

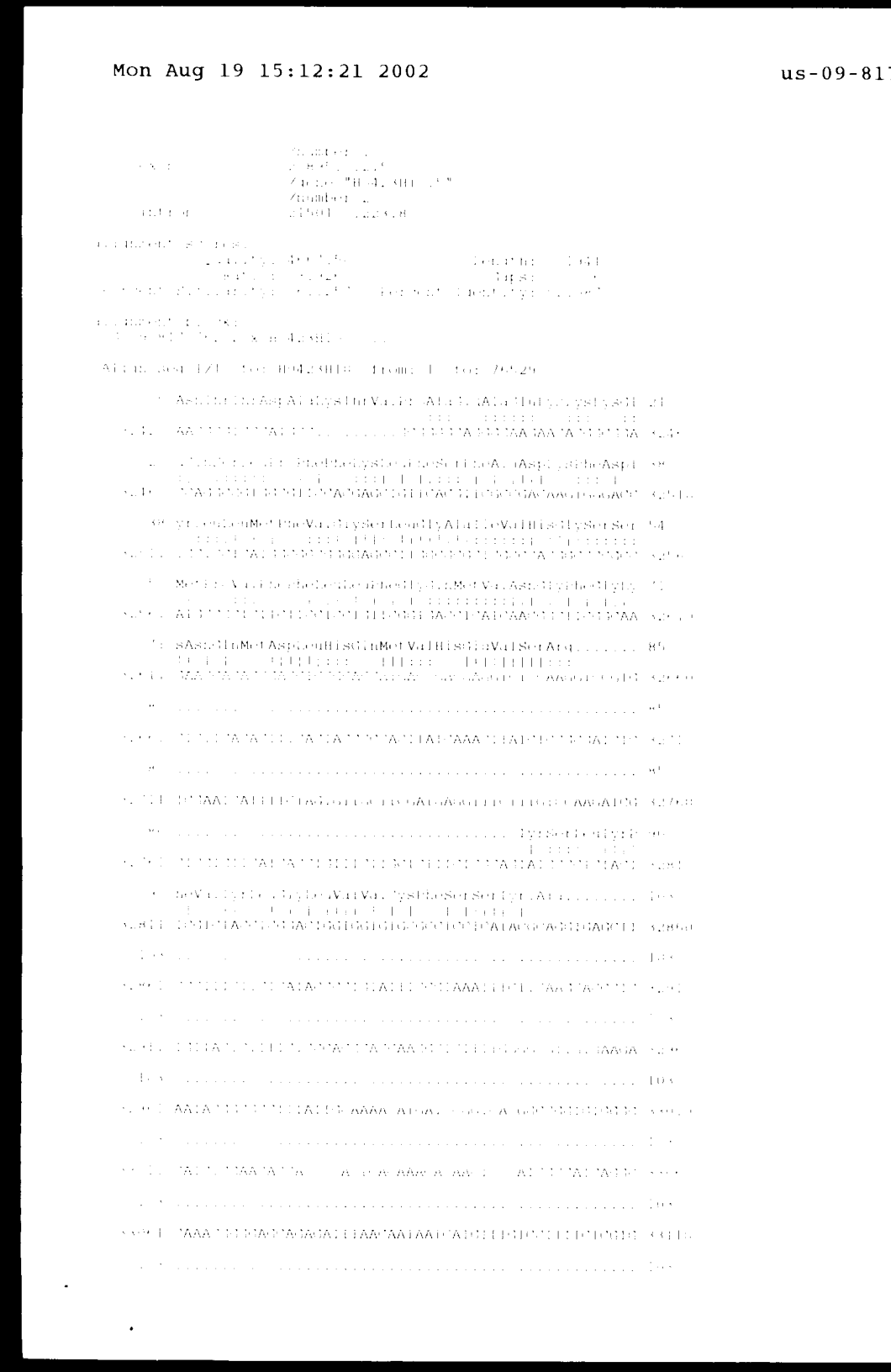
Alignment of: CS 09-817 762-2 2 X SCNBa 9B2/rc to all others
Minimum Match 10%
Shading mode Similarity
Percent Identity 51.4%

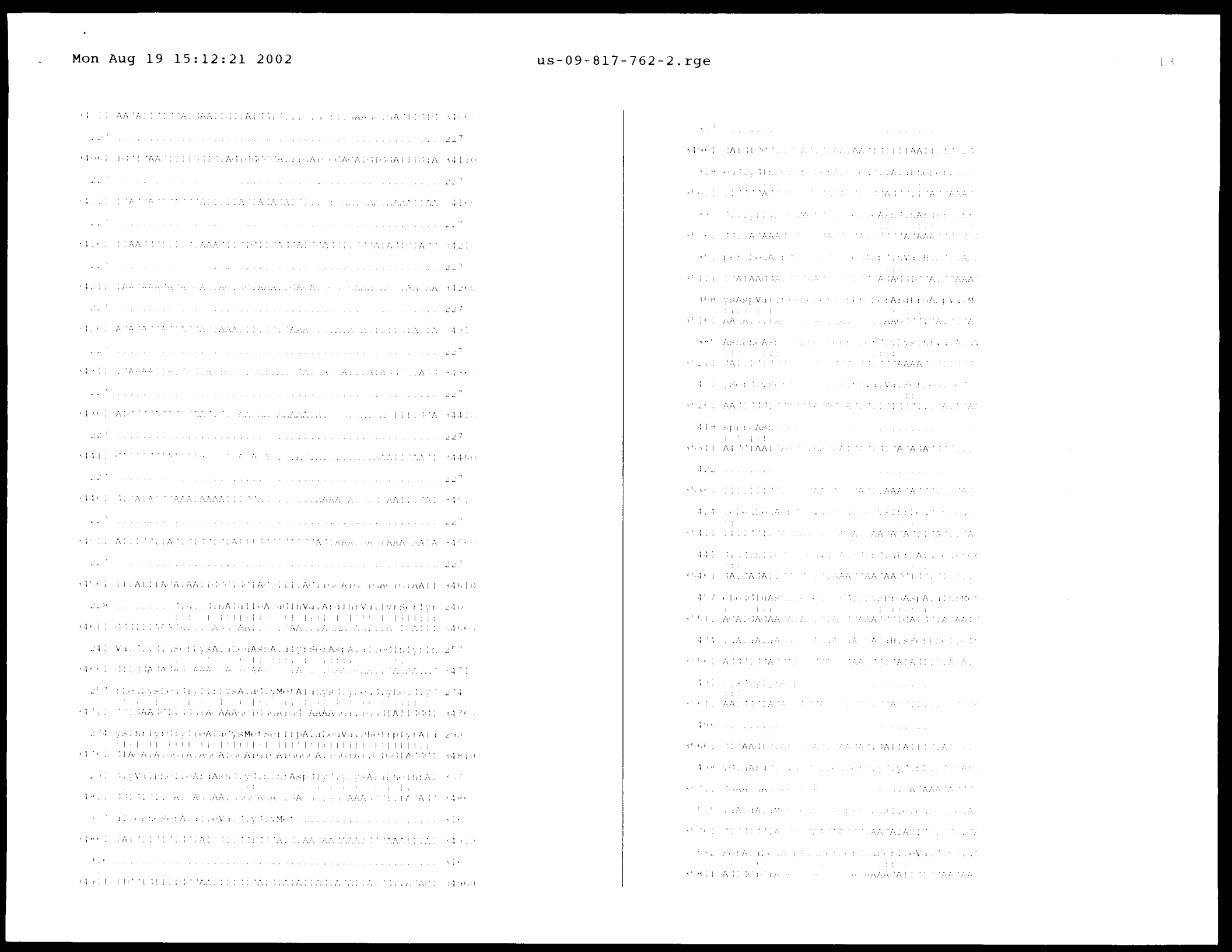


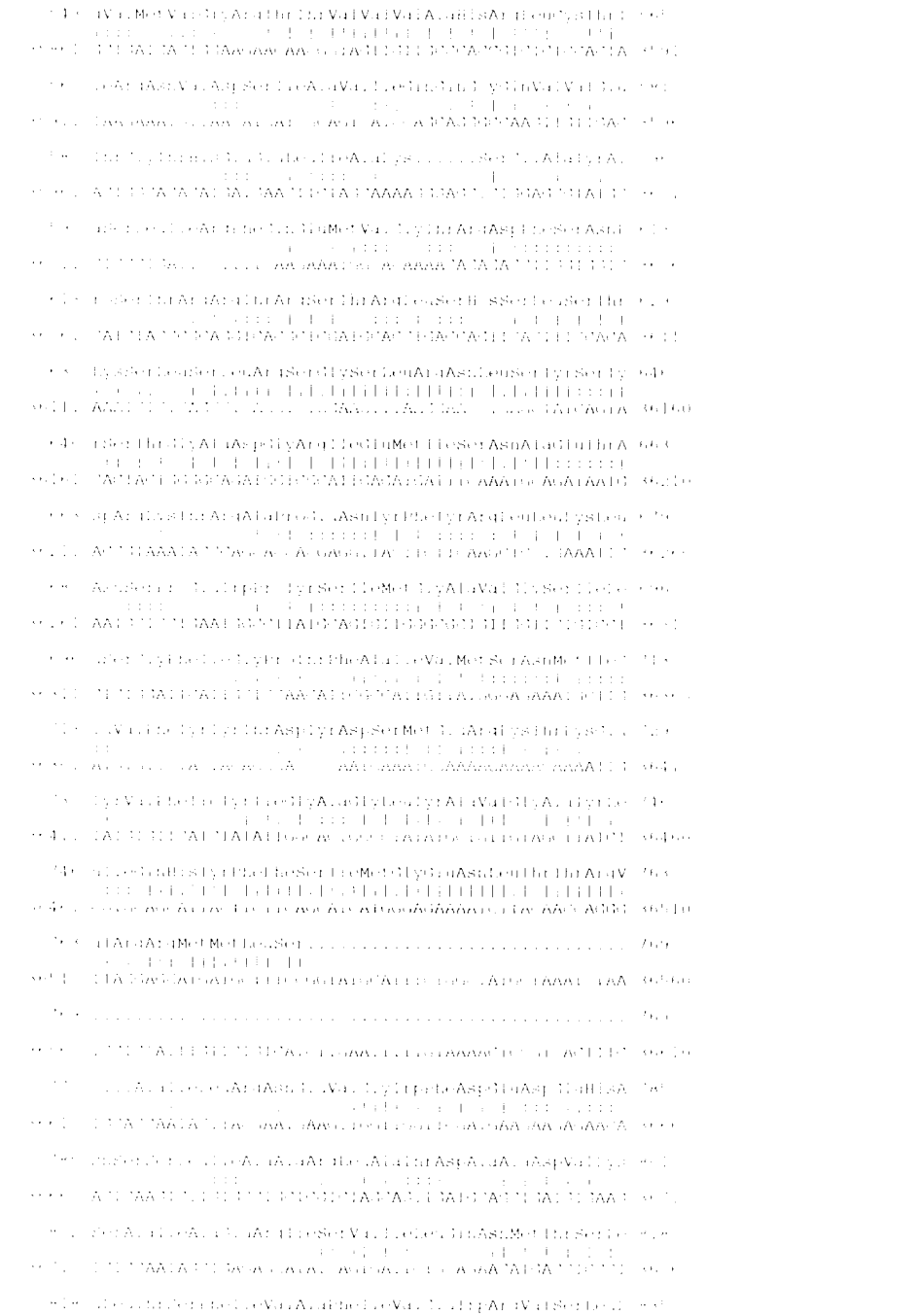












Seq_data: 98,219 SUN: 15

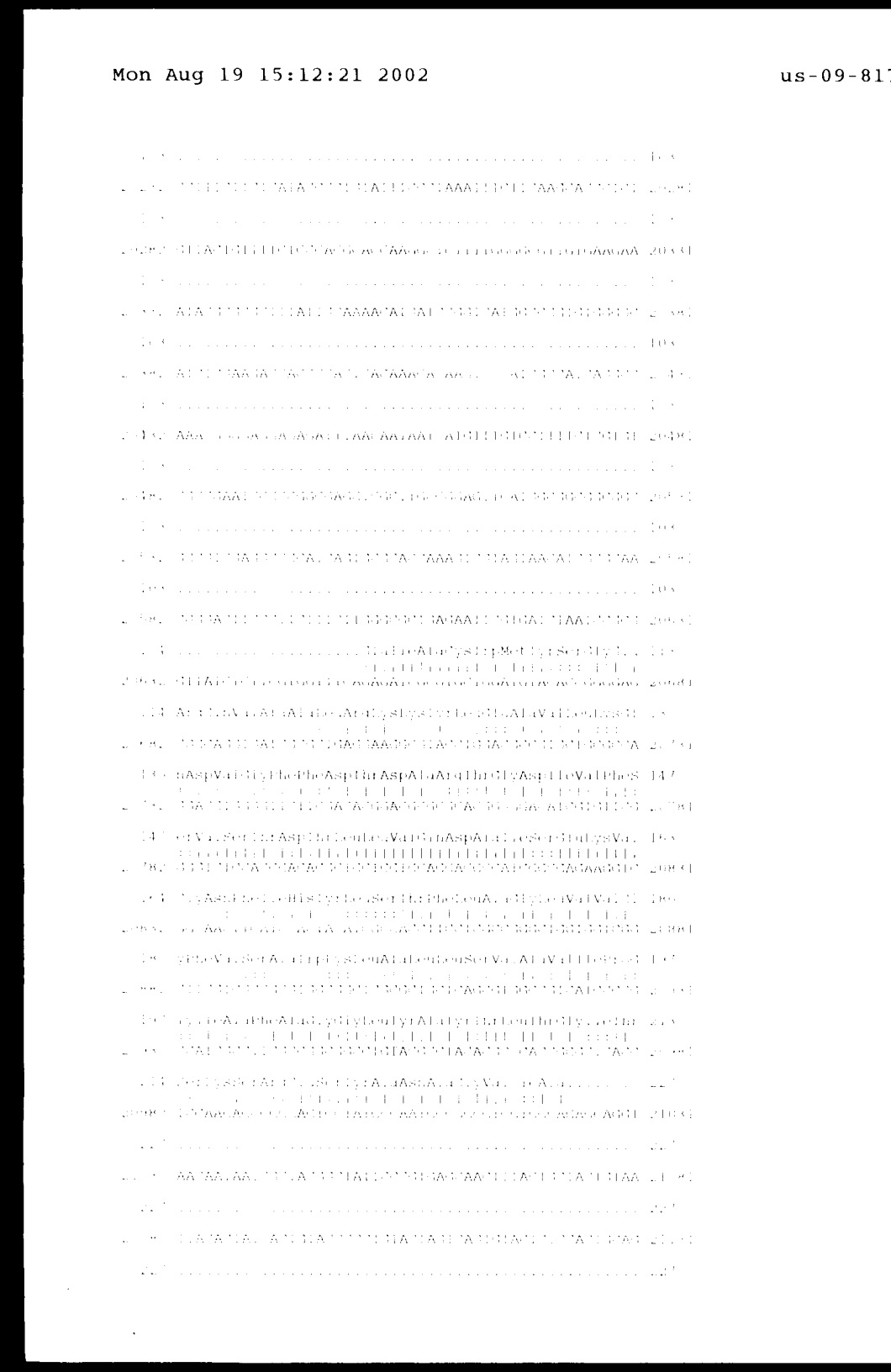
seq_documentation_block:
LOCUS OSUN0015 126786 bp DNA linear HTG 08-SEP-2001
DEFINITION Oryza sativa chromosome 4 clone OSJNBb0072k02, *** SEQUENCING IN PROGRESS ***, 18 ordered pieces.
ACCESSION AL606495
VERSION AL606495.1 GI:19552695
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 126786)
AUTHORS Ha,B., Feng,J., Huang,Y.-..., Li,Y., Zhu,Z.-..., Zhang,..., He,X., Ye,Y.-..., Ma,J., Yu,S.-..., Deng,X.-..., Chen,Y., Wang,..., Zhang,L., Ye,Y., Ye,S., Li,X.-..., Zhang,Y., Li,Y., Li,..., Li,..., Zhang,Y., Hu,H., Jia,P.X., Lan,Y.M., Yin,K., Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,K.-..., Qian,Z.P., Fu,G., Wan,S.Y., Bo,S.X., Li,L., Lu,W., Su,W.-..., Zhu,S.F., Du,Y.F., Ye,..., Yu,H.F., Zhang,Y., Jia,..., Deng,..., Kang,B., Chen,X.Y., Shi,..., Sun,Y., Bao,..., Zhang,X.-..., Zhang,W., Wang,L.-..., Ding,C.W., Shen,H.B., Gu,J.-..., Chen,S.-..., Ni,L., Zhu,F.H. and Huang,J.F.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-...) Bac Sub, National Center for Gene Research,

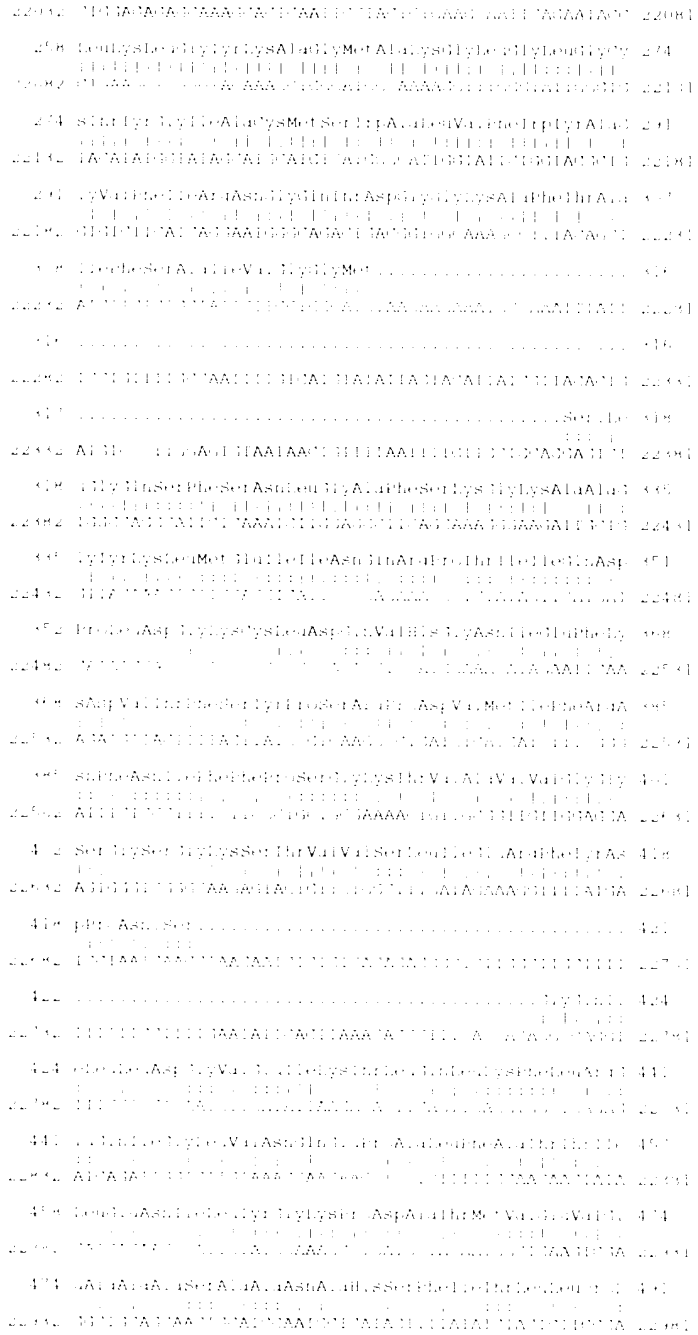
REMARK ...

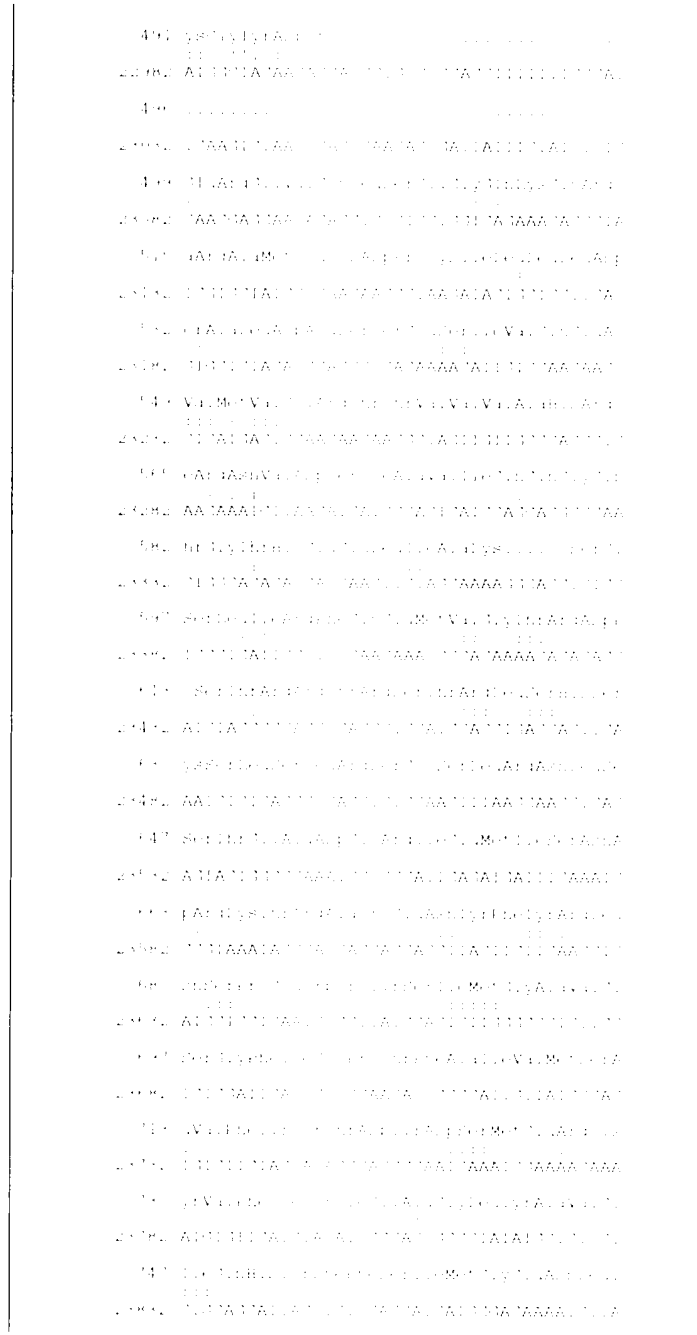
COMMENT We ...
...
...
...
...

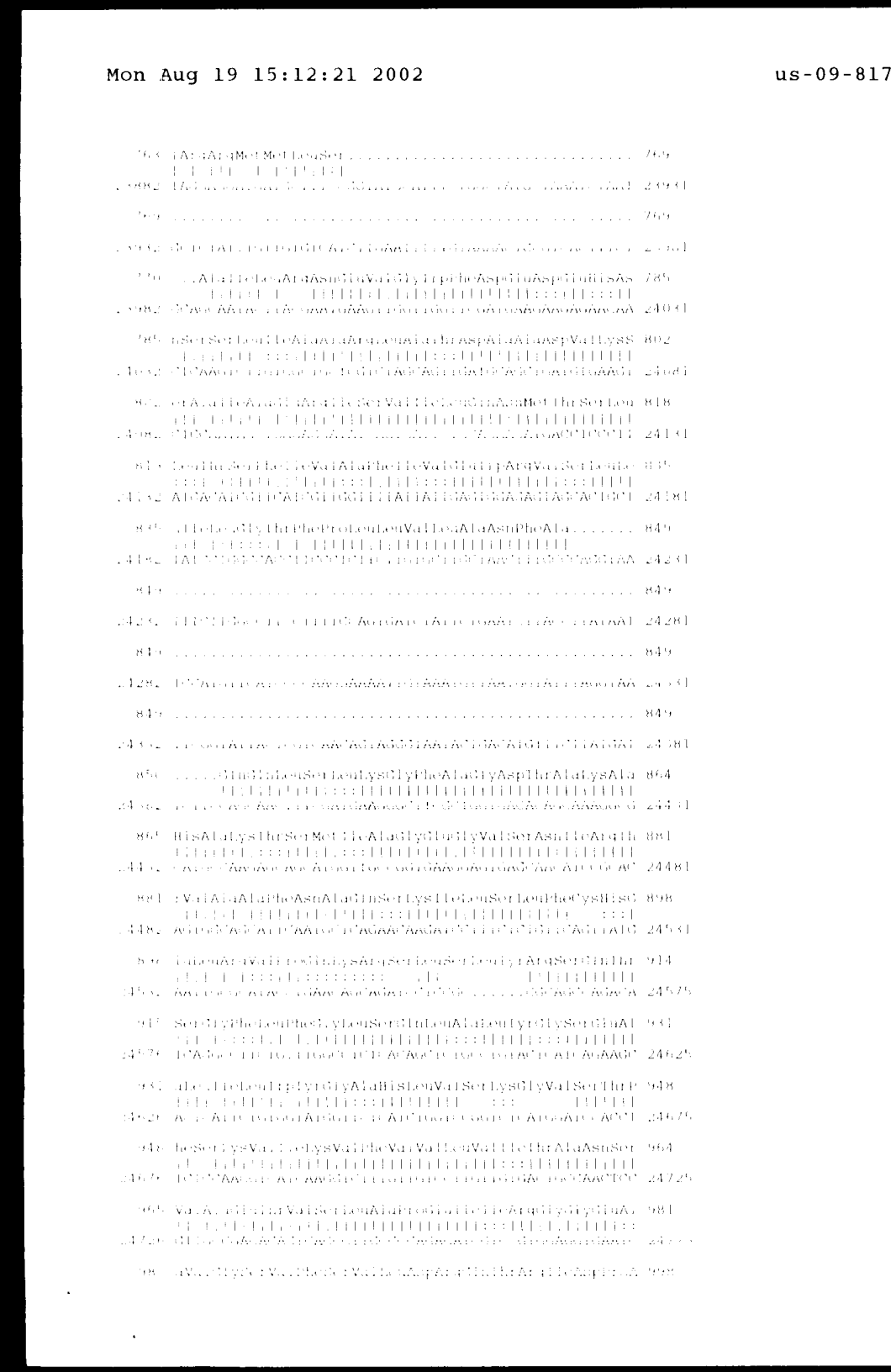
FEATURES ...
source ...
...
...
...
...
...

BASE COUNT ...
ORIGIN



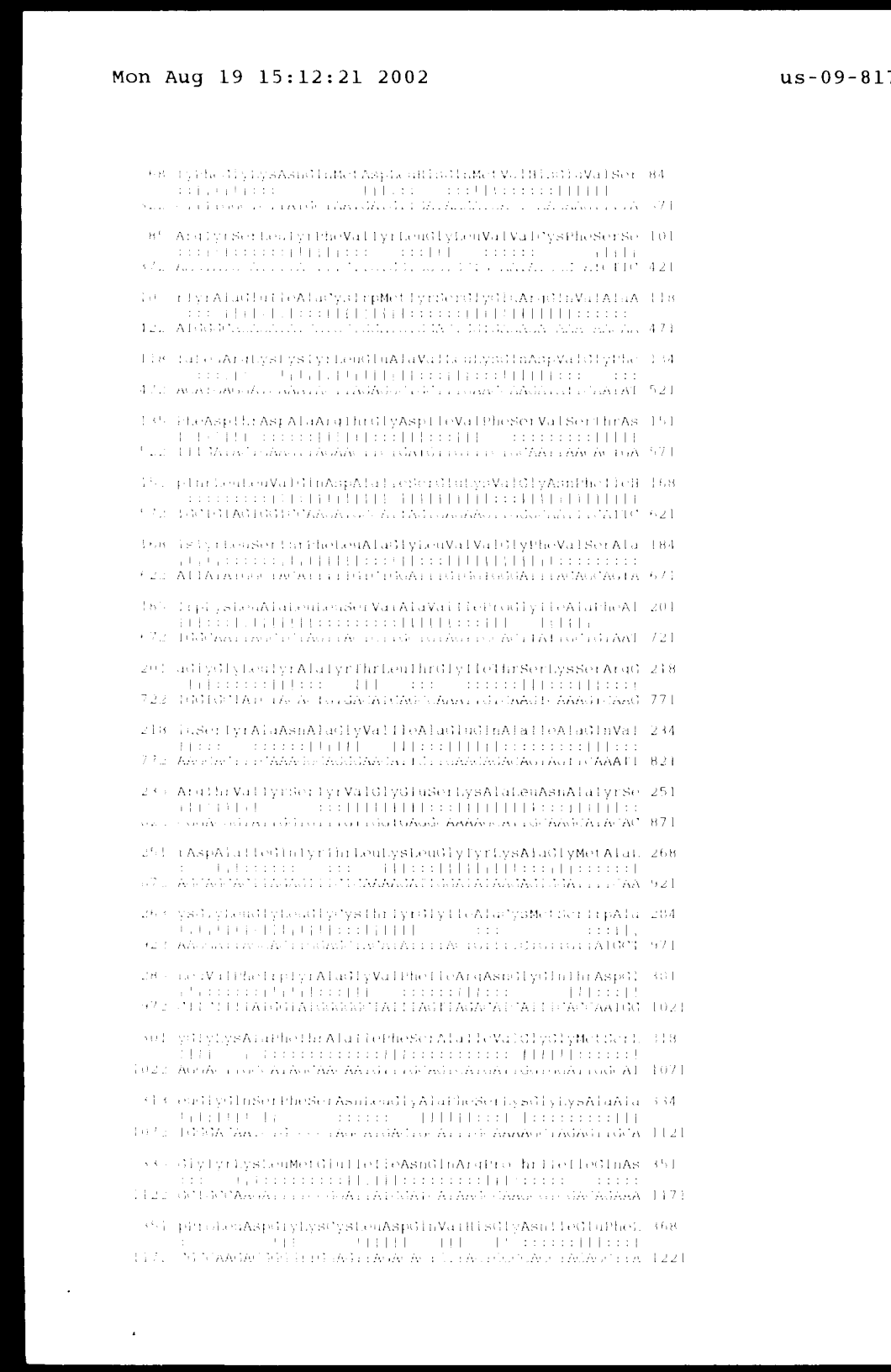


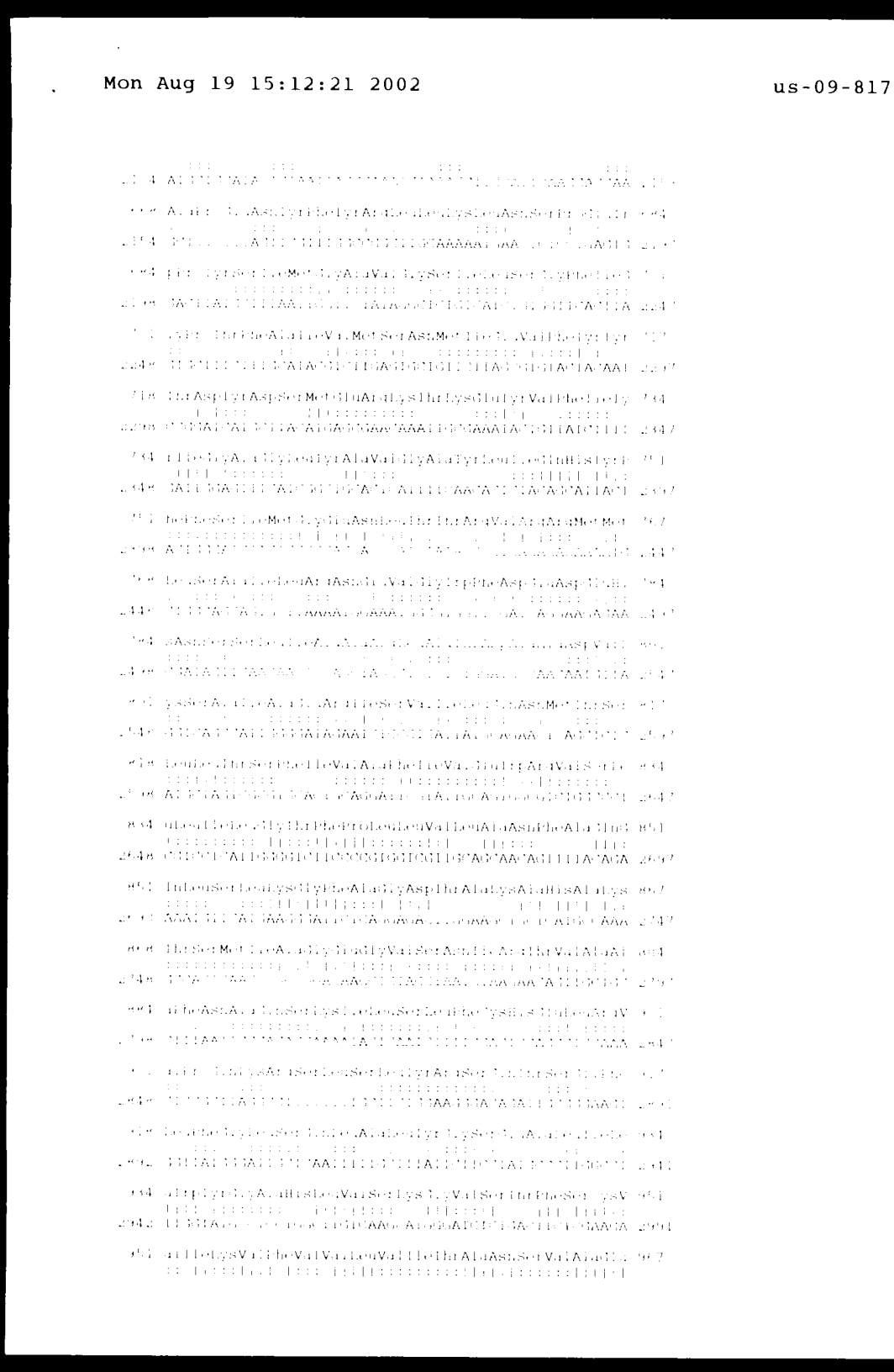


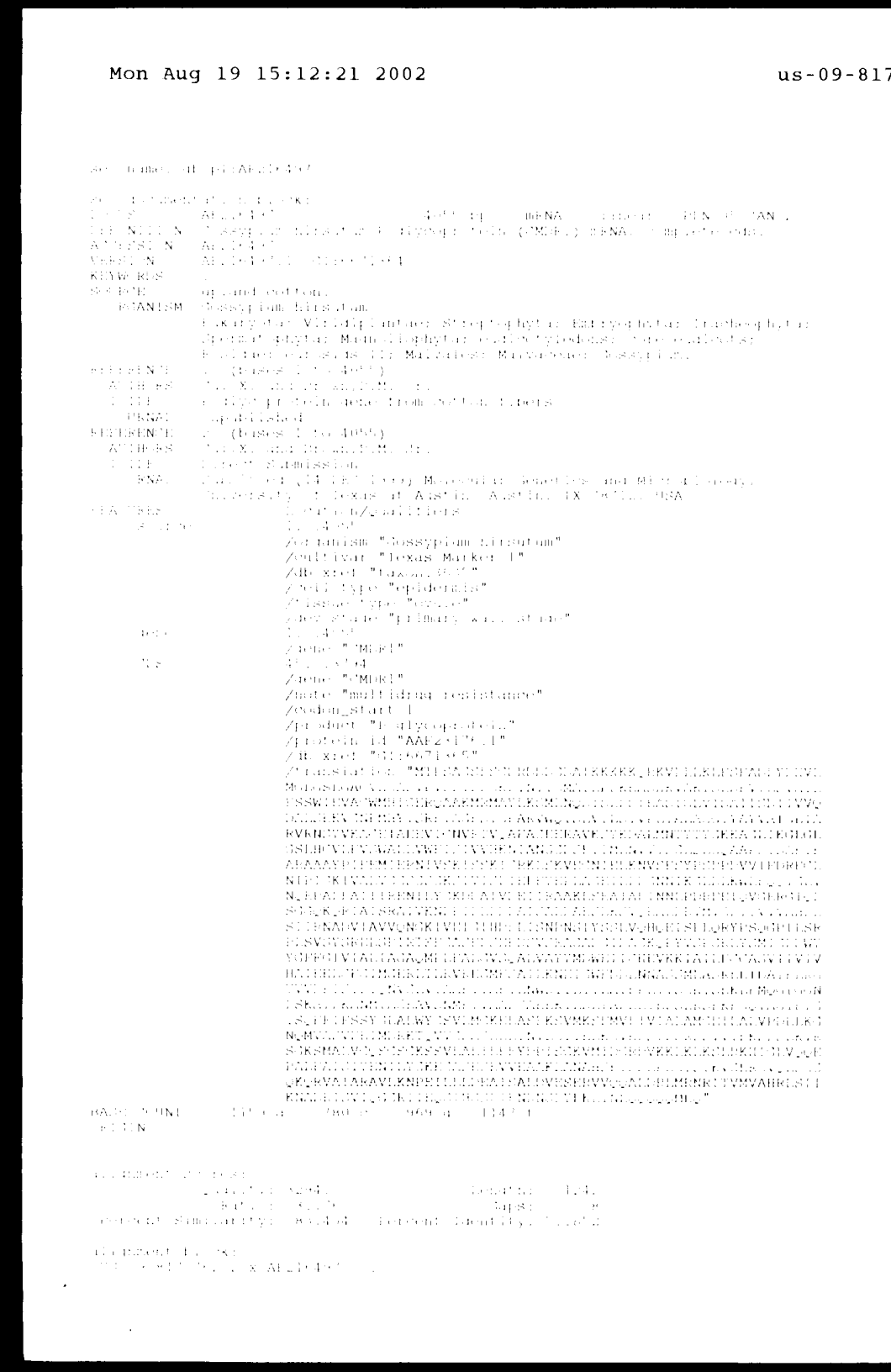


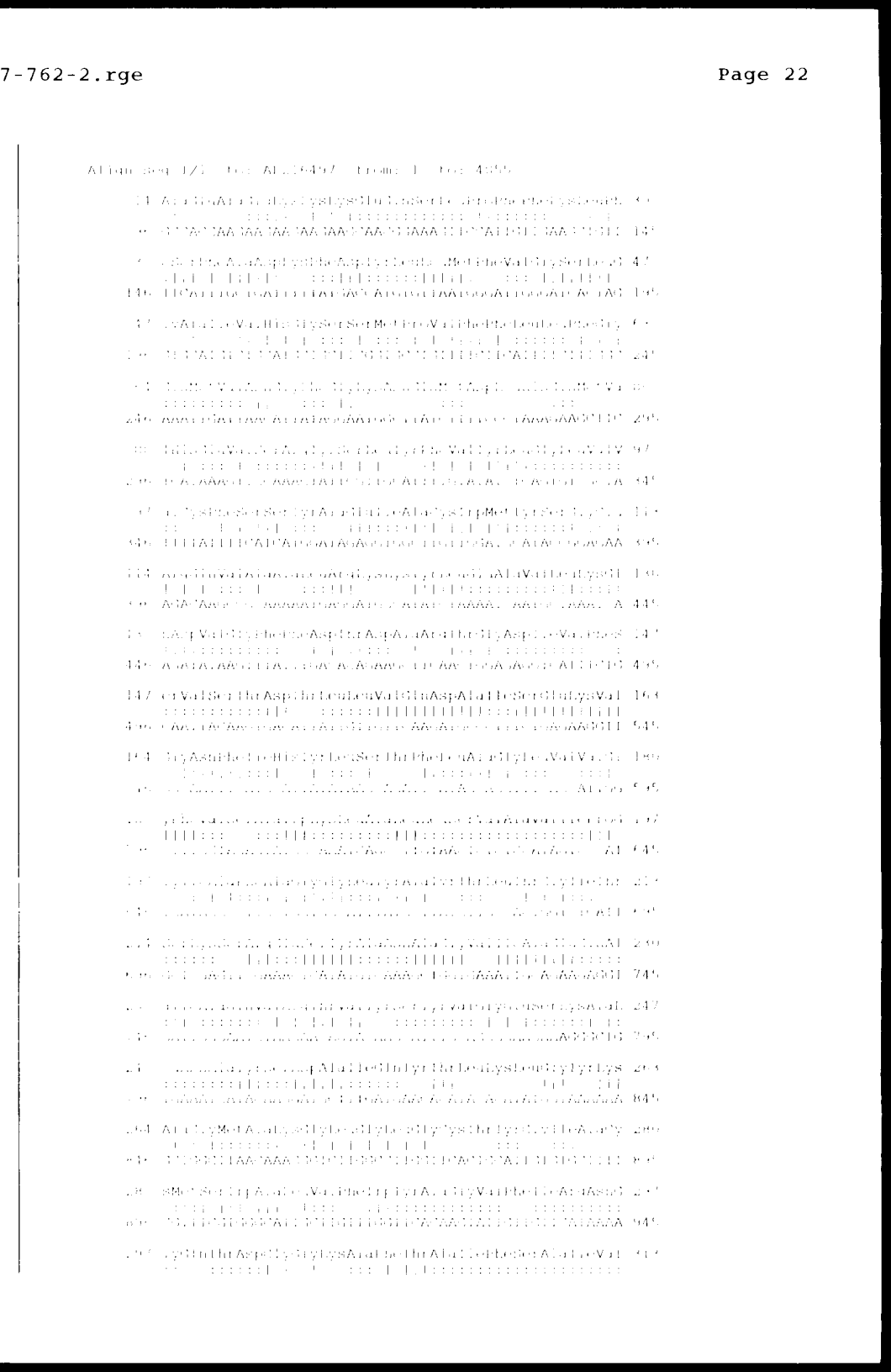
seq_documentation_block:
DEFINITION Solanum tuberosum Proglycoprotein (pmdr1) mRNA, complete cds.
ACCESSION U52079

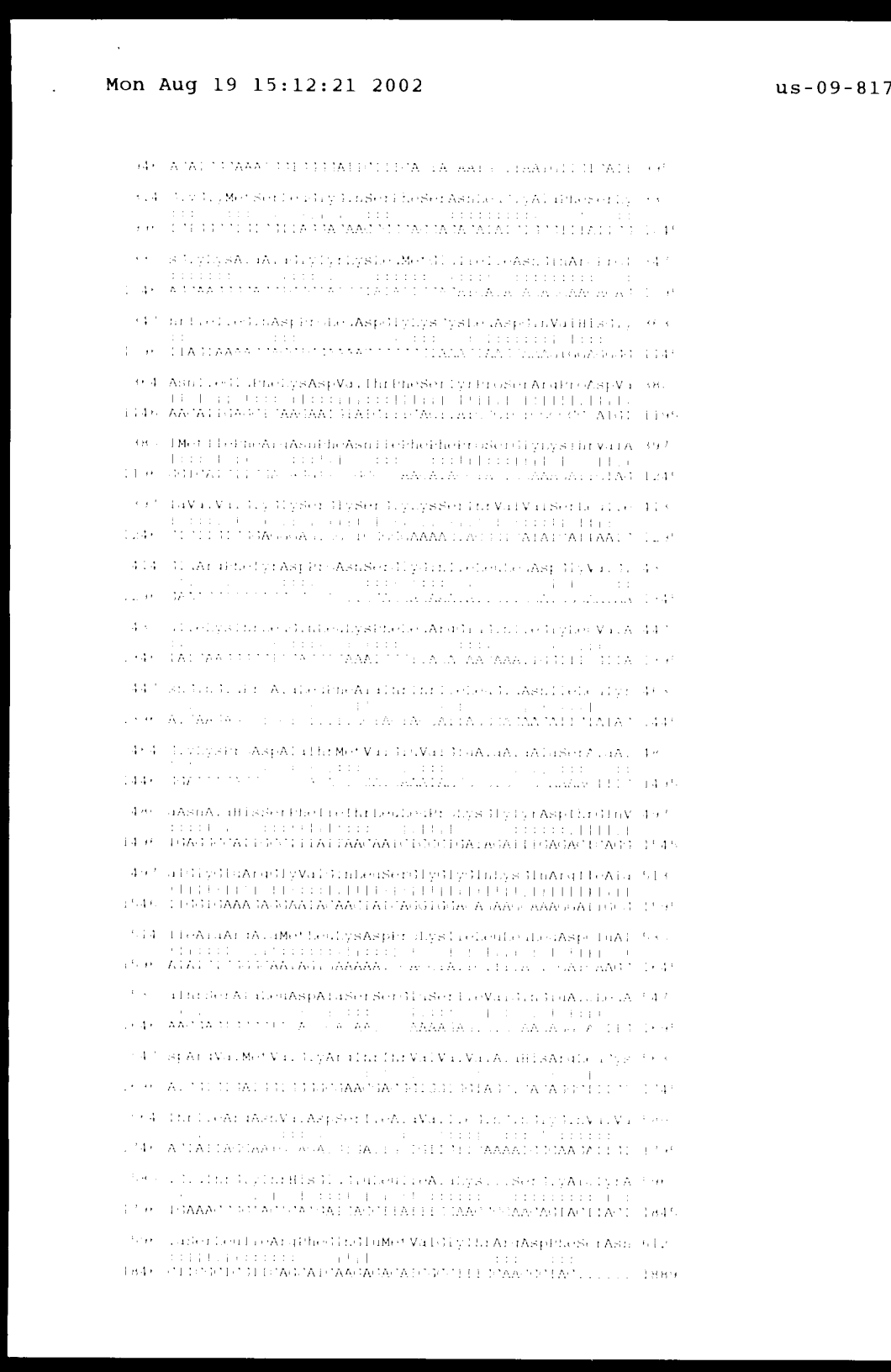
```
VERSION     [illegible]
KEYWORDS
SOURCE      [illegible]
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 4160)
  AUTHORS   Watson,W., Bakeswaite, and Poorstudent,W.
  TITLE     A potato cDNA encoding a homologue of mammalian multidrug resistant
            P-glycoprotein
  JOURNAL   Plant Mol. Biol. 41 (5), 585-597 (1999)
  MEDLINE   [illegible]
   PUBMED   [illegible]
REFERENCE   2  (bases 1 to 4160)
  AUTHORS   Watson,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21 MAR 1999) Wyo. Water Department of Horticulture,
            Westside 5 State Univ.Pullman, WA 99164-6414, USA
FEATURES             Location/Qualifiers
     source          1..4160
                     /chromosome "Solanum tuberosum"
                     /cultivar "Russet Burbank"
                     /db_xref "taxon:4113"
                     /dev_stage "plant"
                     /mol_type "genomic DNA"
                     /tissue_type "stem tuber"
     gene            1..4160
                     /gene "pmdr1"
     CDS             33..3974
                     /gene "pmdr1"
                     /note "binds ATP; ATPases; transporters; transmembrane
                     protein"
                     /codon_start=1
                     /product "P-glycoprotein"
                     /protein_id="AAL[illegible]"
                     /db_xref="GI:14[illegible]"
                     /translation="MGVEEVVVEQPKKNKNGGKKKKNGKLITVTEMFRYADGLDYVLM
                     IVGSIGAIVHGSSMPVFFLFFGKLINIIGLAYLFPAEASHKVAKYALDFVYLGIGIFV
                     CSFLEVACWMHTGERQAARMRLAYLKAVLRQDVAFFDKETNTGEVIGRMSGDTVLIQD
                     AMGEKVGKFIQLVATFFAGFIIAFVKGWLLTLVMLSSIPVLVMAGAVMSIIISKMASR
                     GQIAYAEAGNVVEQTIGSIRTVASFTGEKQAVSRYDKYLVKAYKSGVKEGLVSGLGLG
                     SMFLIIFCSYALAVWYGSKLILEKGYTGGQVINVIIAVLTGSMSLGQASPCIEAFANG
                     RAAAYKMFETIKRKPEIDSYSTEGLKPDKLEGNVEFKDVHFSYPSRPDVQIFRDFSLH
                     VPSGKTVALVGESGSGKSTVISLIERFYDPLAGKVLIDGKEIKRLNVQWLRAHMGIVS
                     QEPILFDCSIAENIAYGDNSRVVSQEEIVAAAKEANIHAFIESLPNKYSTKVGSKGTQ
                     LSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKARVGRTTIVIAHRL
                     STIRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSE
                     IDALEMSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNLVLSIFSLFFIMVATLYAMVKT
                     TRGDGATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFYKQRDLLFYPAWAYAIPSWIL
                     KIPITFIEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLLVNQMASALFRFIGAVGRNMI
                     VANTFGSFALLVIFALGGFILSRDDIKKWWIWGYWISPLMYAQNAISVNEFLGHSWSH
                     VVPNSTESLGVQVLKSRGFFTEAYWYWIGVGALIGFTLLFNILFTLALTYLNPLGKPQ
                     AVISEEELKERDSRRGGEMQELKSGTSHGDSADAERAKMRLPESTSHVDTDIRLLEMP
                     HSEHERSIVQGQVETQDNKGPVQHVVKVSIWWMAYGVHIVIHSLILGFIFQFLQIFSD
                     WWLAYETPADKSKPFNPTVLIEVIYVIVQSFCLLKAMRAVVFVYVTLKASQNLHNKMF
                     ESILRAPMSFFDSTPSGRILNRASTDQSAVDLDIPFQLGFFAFSAIQLLGIIAVMSQV
                     AWQVFIVFIPVIAISIWYQQYYISSARELSRLVGVCKAPVIQHFAESISGAMTIRSFD
                     QTTRFNAALMDHTQDASIFFHSFAMNEWLIQRMETLSAIVLASSLIMVLLPPGTFNPG
                     FVGMALSYGLSLNMVLFWAVYMACFIENKMVSVERIKQFTDIPSEAPWEIKDSAPPPS
                     WPQVGRVEFRDYGLRYREDLDLVLKHINVDIKSGEKVGIVGRTGAGKSSMGVALFRLV
                     EPAAGKIIIDGVDISTIGLHDLRSRLSIIPQDPTMFEGTVRSNLDPLEEHTDEQIWEV
                     LDKCQLGDEVRKKEGKLDSTVVENGENWSVGQRQLVCLGRVLLKRSKILVLDEATASV
                     DTATDNLIQKTIREEFKSCTMLIIAHRLNTIIDSDKIMVLDSGRLKEYDEPYVLLQNK
                     ESLFYKMVQQTGKANAEYLRSLVLGEKEVSKVDEVNSVSSERQKLDKNEAASSLIKKE
                     EQEVPLSRLASLNKPEIPVLLLGSIAAIINGVIFPIFGILISSAIKTFYEPPHELRKD
                     SRFWALMFVVLGVISFITIPARTYFFAVAGCKLIQRIRSMCFEKVVHMEVGWFDEPEH
                     SSGAIGARLSADAASVRSLVGDALALIVQNISTAVTGLIIAFTANWRLALIVIAVLPL
                     IIGSTMQMKMLSGFAADAKAMYEEASQVANDAVGSIRTVASFCAEEKVMELYKKKCEG
                     PMKTGIRQGLISGIGFGFSFFLIFSSYALALWYGAKLILDKGYTFRAVINVFMAILTG
                     AMALGQASSFSPDYAKAKISAAHIFMIIERQPLIDSYSEEGLMPNTLEGNVTFGEVVF
                     NYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGMVSIDGQE
                     IRSLNVQWLRAQMGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVAAAKEANIHAFI
                     ESLPNKYSTKVGSKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQE
                     ALDKARVGRTTIVIAHRLSTIRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQ
                     TAGNEVELENAADESKSEIDALEMSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNLVLS
                     IFSLFFIMVATLYAMVKTTRGDGATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFYKQR
                     DLLFYPAWAYAIPSWILKIPITFIEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLLVNQ
                     MASALFRFIGAVGRNMIVANTFGSFALLVIFALGGFILSRDDIKKWWIWGYWISPLMY
                     AQNAISVNEFLGHSWSHVVPNSTESLGVQVLKSRGFFTEAYWYWIGVGALIGFTLLFN
                     ILFTLALTYLNPLGKPQAVISEEELKERDSRRGGEMQELKSGTSHGDSADAERAKMRL
                     PESTSHVDTDIRLLEMPHSEHERSIVQGQVETQDNKGPVQHVVKVSIWWMAYGVHIVI
                     HSLILGFIFQFLQIFSDWWLAYETPADKSKPFNPTVLIEVIYVIVQSFCLLKAMRAVV
                     FVYVTLKASQNLHNKMFESILRAPMSFFDSTPSGRILNRASTDQSAVDLDIPFQLGFF
                     AFSAIQLLGIIAVMSQVAWQVFIVFIPVIAISIWYQQYYISSARELSRLVGVCKAPVI
                     QHFAESISGAMTIRSFDQTTRFNAALMDHTQDASIFFHSFAMNEWLIQRMETLSAIVL
                     ASSLIMVLLPPGTFNPGFVGMALSYGLSLNMVLFWAVYMACFIENKMVSVERIKQFTD
                     IPSEAPWEIKDSAPPPSWPQVGRVEFRDYGLRYREDLDLVLKHINVDIKSGEKVGIVG
                     RTGAGKSSMGVALFRLVEPAAGKIIIDGVDISTIGLHDLRSRLSIIPQDPTMFEGTVR
                     SNLDPLEEHTDEQIWEVLDKCQLGDEVRKKEGKLDSTVVENGENWSVGQRQLVCLGRV
                     LLKRSKILVLDEATASVDTATDNLIQKTIREEFKSCTMLIIAHRLNTIIDSDKIMVLD
                     SGRLKEYDEPYVLLQNKESLFYKMVQQTGKANAEYLRSLVLGEKEVSKVDEVNSVSSE
                     RQKLDKNEAASSLIKKEEQEVPLSRLASLNKPEIPVLLLGSIAAIINGVIFPIFGILI
                     SSAIKTFYEPPHELRKDSRFWALMFVVLGVISFITIPARTYFFAVAGCKLIQRIRSMC
                     FEKVVHMEVGWFDEPEHSSGAIGARLSADAASVRSLVGDALALIVQNISTAVTGLIIA
                     FTANWRLALIVIAVLPLIIGSTMQMKMLSGFAADAKAMYEEASQVANDAVGSIRTVAS
                     FCAEEKVMELYKKKCEGPMKTGIRQGLISGIGFGFSFFLIFSSYALALWYGAKLILDK
                     GYTFRAVINVFMAILTGAMALGQASSFSPDYAKAKISAAHIFMIIERQPLIDSYSEEG
                     LMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLL
                     ERFYDPLAGMVSIDGQEIRSLNVQWLRAQMGIVSQEPILFDCSIAENIAYGDNSRVVS
                     QEEIVAAAKEANIHAFIESLPNKYSTKVGSKGTQLSGGQKQRIAIARALVRQPHILLL
                     DEATSALDTESEKVVQEALDKARVGRTTIVIAHRLSTIRNADVIAGFDDGVIVEKGNH
                     DELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDYRSSLIRAIYRSL
                     GRSFVYVFRGTSNLVLSIFSLFFIMVATLYAMVKTTRGDGATVMGSLYAVTMIIVSGY
                     VEVSMTLAKLPVFYKQRDLLFYPAWAYAIPSWILKIPITFIEVAIWVFLTYYVIGFDP
                     NVGRLFKQYLLLLLVNQMASALFRFIGAVGRNMIVANTFGSFALLVIFALGGFILSRD
                     DIKKWWIWGYWISPLMYAQNAISVNEFLGHSWSHVVPNSTESLGVQVLKSRGFFTEAY
                     WYWIGVGALIGFTLLFNILFTLALTYLNPLGKPQAVISEEELKERDSRRGGEMQELKS
                     GTSHGDSADAERAKMRLPESTSHVDTDIRLLEMPHSEHERSIVQGQVETQDNKGPVQH
                     VVKVSIWWMAYGVHIVIHSLILGFIFQFLQIFSDWWLAYETPADKSKPFNPTVLIEVI
                     YVIVQSFCLLKAMRAVVFVYVTLKASQNLHNKMFESILRAPMSFFDSTPSGRILNRAS
                     TDQSAVDLDIPFQLGFFAFSAIQLLGIIAVMSQVAWQVFIVFIPVIAISIWYQQYYIS
                     SARELSRLVGVCKAPVIQHFAESISGAMTIRSFDQTTRFNAALMDHTQDASIFFHSFA
                     MNEWLIQRMETLSAIVLASSLIMVLLPPGTFNPGFVGMALSYGLSLNMVLFWAVYMAC
                     FIENKMVSVERIKQFTDIPSEAPWEIKDSAPPPSWPQVGRVEFRDYGLRYREDLDLVL
                     KHINVDIKSGEKVGIVGRTGAGKSSMGVALFRLVEPAAGKIIIDGVDISTIGLHDLRS
                     RLSIIPQDPTMFEGTVRSNLDPLEEHTDEQIWEVLDKCQLGDEVRKKEGKLDSTVVE
                     NGENWSVGQRQLVCLGRVLLKRSKILVLDEATASVDTATDNLIQKTIREEFKSCTMLI
                     IAHRLNTIIDSDKIMVLDSGRLKEYDEPYVLLQNKESLFYKMVQQTGKANAEYLRSLV
                     LGEKEVSKVDEVNSVSSERQKLDKNEAASSLIKKEEQEVPLSRLASLNKPEIPVLLLG
                     SIAAIINGVIFPIFGILISSAIKTFYEPPHELRKDSRFWALMFVVLGVISFITIPART
                     YFFAVAGCKLIQRIRSMCFEKVVHMEVGWFDEPEHSSGAIGARLSADAASVRSLVGDA
                     LALIVQNISTAVTGLIIAFTANWRLALIVIAVLPLIIGSTMQMKMLSGFAADAKAMYE
                     EASQVANDAVGSIRTVASFCAEEKVMELYKKKCEGPMKTGIRQGLISGIGFGFSFFLI
                     FSSYALALWYGAKLILDKGYTFRAVINVFMAILTGAMALGQASSFSPDYAKAKISAAH
                     IFMIIERQPLIDSYSEEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQT
                     LALVGSSGCGKSTVVQLLERFYDPLAGMVSIDGQEIRSLNVQWLRAQMGIVSQEPILF
                     DCSIAENIAYGDNSRVVSQEEIVAAAKEANIHAFIESLPNKYSTKVGSKGTQLSGGQK
                     QRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKARVGRTTIVIAHRLSTIRN
                     ADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALE
                     MSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNLVLSIFSLFFIMVATLYAMVKTTRGDG
                     ATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFYKQRDLLFYPAWAYAIPSWILKIPITF
                     IEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLLVNQMASALFRFIGAVGRNMIVANTFG
                     SFALLVIFALGGFILSRDDIKKWWIWGYWISPLMYAQNAISVNEFLGHSWSHVVPNST
                     ESLGVQVLKSRGFFTEAYWYWIGVGALIGFTLLFNILFTLALTYLNPLGKPQAVISEE
                     ELKERDSRRGGEMQELKSGTSHGDSADAERAKMRLPESTSHVDTDIRLLEMPHSEHER
                     SIVQGQVETQDNKGPVQHVVKVSIWWMAYGVHIVIHSLILGFIFQFLQIFSDWWLAYE
                     TPADKSKPFNPTVLIEVIYVIVQSFCLLKAMRAVVFVYVTLKASQNLHNKMFESILRA
                     PMSFFDSTPSGRILNRASTDQSAVDLDIPFQLGFFAFSAIQLLGIIAVMSQVAWQVFI
                     VFIPVIAISIWYQQYYISSARELSRLVGVCKAPVIQHFAESISGAMTIRSFDQTTRFN
                     AALMDHTQDASIFFHSFAMNEWLIQRMETLSAIVLASSLIMVLLPPGTFNPGFVGMAL
                     SYGLSLNMVLFWAVYMACFIENKMVSVERIKQFTDIPSEAPWEIKDSAPPPSWPQVGR
                     VEFRDYGLRYREDLDLVLKHINVDIKSGEKVGIVGRTGAGKSSMGVALFRLVEPAAGK
                     IIIDGVDISTIGLHDLRSRLSIIPQDPTMFEGTVRSNLDPLEEHTDEQIWEVLDKCQL
                     GDEVRKKEGKLDSTVVENGENWSVGQRQLVCLGRVLLKRSKILVLDEATASVDTATDN
                     LIQKTIREEFKSCTMLIIAHRLNTIIDSDKIMVLDSGRLKEYDEPYVLLQNKESLFYK
                     MVQQTGKANAEYLRSLVLGEKEVSKVDEVNSVSSERQKLDKNEAASSLIKKEEQEVPL
                     SRLASLNKPEIPVLLLGSIAAIINGVIFPIFGILISSAIKTFYEPPHELRKDSRFWAL
                     MFVVLGVISFITIPARTYFFAVAGCKLIQRIRSMCFEKVVHMEVGWFDEPEHSSGAIG
                     ARLSADAASVRSLVGDALALIVQNISTAVTGLIIAFTANWRLALIVIAVLPLIIGSTM
                     QMKMLSGFAADAKAMYEEASQVANDAVGSIRTVASFCAEEKVMELYKKKCEGPMKTGI
                     RQGLISGIGFGFSFFLIFSSYALALWYGAKLILDKGYTFRAVINVFMAILTGAMALGQ
                     ASSFSPDYAKAKISAAHIFMIIERQPLIDSYSEEGLMPNTLEGNVTFGEVVFNYPTRP
                     DIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGMVSIDGQEIRSLNV
                     QWLRAQMGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVAAAKEANIHAFIESLPNK
                     YSTKVGSKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKAR
                     VGRTTIVIAHRLSTIRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEV
                     ELENAADESKSEIDALEMSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNLVLSIFSLFF
                     IMVATLYAMVKTTRGDGATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFYKQRDLLFYP
                     AWAYAIPSWILKIPITFIEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLLVNQMASALF
                     RFIGAVGRNMIVANTFGSFALLVIFALGGFILSRDDIKKWWIWGYWISPLMYAQNAIS
                     VNEFLGHSWSHVVPNSTESLGVQVLKSRGFFTEAYWYWIGVGALIGFTLLFNILFTLA
                     LTYLNPLGKPQAVISEEELKERDSRRGGEMQELKSGTSHGDSADAERAKMRLPESTSH
                     VDTDIRLLEMPHSEHERSIVQGQVETQDNKGPVQHVVKVSIWWMAYGVHIVIHSLILG
                     FIFQFLQIFSDWWLAYETPADKSKPFNPTVLIEVIYVIVQSFCLLKAMRAVVFVYVTL
                     KASQNLHNKMFESILRAPMSFFDSTPSGRILNRASTDQSAVDLDIPFQLGFFAFSAIQ
                     LLGIIAVMSQVAWQVFIVFIPVIAISIWYQQYYISSARELSRLVGVCKAPVIQHFAES
                     ISGAMTIRSFDQTTRFNAALMDHTQDASIFFHSFAMNEWLIQRMETLSAIVLASSLIM
                     VLLPPGTFNPGFVGMALSYGLSLNMVLFWAVYMACFIENKMVSVERIKQFTDIPSEAP
                     WEIKDSAPPPSWPQVGRVEFRDYGLRYREDLDLVLKHINVDIKSGEKVGIVGRTGAGK
                     SSMGVALFRLVEPAAGKIIIDGVDISTIGLHDLRSRLSIIPQDPTMFEGTVRSNLDPL
                     EEHTDEQIWEVLDKCQLGDEVRKKEGKLDSTVVENGENWSVGQRQLVCLGRVLLKRSK
                     ILVLDEATASVDTATDNLIQKTIREEFKSCTMLIIAHRLNTIIDSDKIMVLDSGRLKE
                     YDEPYVLLQNKESLFYKMVQQTGKANAEYLRSLVLGEKEVSKVDEVNSVSSERQKLDK
                     NEAASSLIKKEEQEVPLSRLASLNKPEIPVLLLGSIAAIINGVIFPIFGILISSAIKT
                     FYEPPHELRKDSRFWALMFVVLGVISFITIPARTYFFAVAGCKLIQRIRSMCFEKVVH
                     MEVGWFDEPEHSSGAIGARLSADAASVRSLVGDALALIVQNISTAVTGLIIAFTANWR
                     LALIVIAVLPLIIGSTMQMKMLSGFAADAKAMYEEASQVANDAVGSIRTVASFCAEEK
                     VMELYKKKCEGPMKTGIRQGLISGIGFGFSFFLIFSSYALALWYGAKLILDKGYTFRA
                     VINVFMAILTGAMALGQASSFSPDYAKAKISAAHIFMIIERQPLIDSYSEEGLMPNTL
                     EGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDP
                     LAGMVSIDGQEIRSLNVQWLRAQMGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVA
                     AAKEANIHAFIESLPNKYSTKVGSKGTQLSGGQKQRIAIARALVRQPHILLLDEATSA
                     LDTESEKVVQEALDKARVGRTTIVIAHRLSTIRNADVIAGFDDGVIVEKGNHDELMKE
                     KGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDYRSSLIRAIYRSLGRSFVY
                     VFRGTSNLVLSIFSLFFIMVATLYAMVKTTRGDGATVMGSLYAVTMIIVSGYVEVSMT
                     LAKLPVFYKQRDLLFYPAWAYAIPSWILKIPITFIEVAIWVFLTYYVIGFDPNVGRLF
                     KQYLLLLLVNQMASALFRFIGAVGRNMIVANTFGSFALLVIFALGGFILSRDDIKKWW
                     IWGYWISPLMYAQNAISVNEFLGHSWSHVVPNSTESLGVQVLKSRGFFTEAYWYWIGV
                     GALIGFTLLFNILFTLALTYLNPLGKPQAVISEEELKERDSRRGGEMQELKSGTSHGD
                     SADAERAKMRLPESTSHVDTDIRLLEMPHSEHERSIVQGQVETQDNKGPVQHVVKVSI
                     WWMAYGVHIVIHSLILGFIFQFLQIFSDWWLAYETPADKSKPFNPTVLIEVIYVIVQS
                     FCLLKAMRAVVFVYVTLKASQNLHNKMFESILRAPMSFFDSTPSGRILNRASTDQSAV
                     DLDIPFQLGFFAFSAIQLLGIIAVMSQVAWQVFIVFIPVIAISIWYQQYYISSARELS
                     RLVGVCKAPVIQHFAESISGAMTIRSFDQTTRFNAALMDHTQDASIFFHSFAMNEWLI
                     QRMETLSAIVLASSLIMVLLPPGTFNPGFVGMALSYGLSLNMVLFWAVYMACFIENKM
                     VSVERIKQFTDIPSEAPWEIKDSAPPPSWPQVGRVEFRDYGLRYREDLDLVLKHINVD
                     IKSGEKVGIVGRTGAGKSSMGVALFRLVEPAAGKIIIDGVDISTIGLHDLRSRLSIIP
                     QDPTMFEGTVRSNLDPLEEHTDEQIWEVLDKCQLGDEVRKKEGKLDSTVVENGENWS
                     VGQRQLVCLGRVLLKRSKILVLDEATASVDTATDNLIQKTIREEFKSCTMLIIAHRLN
                     TIIDSDKIMVLDSGRLKEYDEPYVLLQNKESLFYKMVQQTGKANAEYLRSLVLGEKEV
                     SKVDEVNSVSSERQKLDKNEAASSLIKKEEQEVPLSRLASLNKPEIPVLLLGSIAAII
                     NGVIFPIFGILISSAIKTFYEPPHELRKDSRFWALMFVVLGVISFITIPARTYFFAVA
                     GCKLIQRIRSMCFEKVVHMEVGWFDEPEHSSGAIGARLSADAASVRSLVGDALALIVQ
                     NISTAVTGLIIAFTANWRLALIVIAVLPLIIGSTMQMKMLSGFAADAKAMYEEASQVA
                     NDAVGSIRTVASFCAEEKVMELYKKKCEGPMKTGIRQGLISGIGFGFSFFLIFSSYAL
                     ALWYGAKLILDKGYTFRAVINVFMAILTGAMALGQASSFSPDYAKAKISAAHIFMIIE
                     RQPLIDSYSEEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGS
                     SGCGKSTVVQLLERFYDPLAGMVSIDGQEIRSLNVQWLRAQMGIVSQEPILFDCSIAE
                     NIAYGDNSRVVSQEEIVAAAKEANIHAFIESLPNKYSTKVGSKGTQLSGGQKQRIAIA
                     RALVRQPHILLLDEATSALDTESEKVVQEALDKARVGRTTIVIAHRLSTIRNADVIAG
                     FDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDY
                     RSSLIRAIYRSLGRSFVYVFRGTSNLVLSIFSLFFIMVATLYAMVKTTRGDGATVMGS
                     LYAVTMIIVSGYVEVSMTLAKLPVFYKQRDLLFYPAWAYAIPSWILKIPITFIEVAIW
                     VFLTYYVIGFDPNVGRLFKQYLLLLLVNQMASALFRFIGAVGRNMIVANTFGSFALLV
                     IFALGGFILSRDDIKKWWIWGYWISPLMYAQNAISVNEFLGHSWSHVVPNSTESLGVQ
                     VLKSRGFFTEAYWYWIGVGALIGFTLLFNILFTLALTYLNPLGKPQAVISEEELKERD
                     SRRGGEMQELKSGTSHGDSADAERAKMRLPESTSHVDTDIRLLEMPHSEHERSIVQGQ
                     VETQDNKGPVQHVVKVSIWWMAYGVHIVIHSLILGFIFQFLQIFSDWWLAYETPADKS
                     KPFNPTVLIEVIYVIVQSFCLLKAMRAVVFVYVTLKASQNLHNKMFESILRAPMSFFD
                     STPSGRILNRASTDQSAVDLDIPFQLGFFAFSAIQLLGIIAVMSQVAWQVFIVFIPVI
                     AISIWYQQYYISSARELSRLVGVCKAPVIQHFAESISGAMTIRSFDQTTRFNAALMDH
                     TQDASIFFHSFAMNEWLIQRMETLSAIVLASSLIMVLLPPGTFNPGFVGMALSYGLSL
                     NMVLFWAVYMACFIENKMVSVERIKQFTDIPSEAPWEIKDSAPPPSWPQVGRVEFRDY
                     GLRYREDLDLVLKHINVDIKSGEKVGIVGRTGAGKSSMGVALFRLVEPAAGKIIIDGV
                     DISTIGLHDLRSRLSIIPQDPTMFEGTVRSNLDPLEEHTDEQIWEVLDKCQLGDEVRK
                     KEGKLDSTVVENGENWSVGQRQLVCLGRVLLKRSKILVLDEATASVDTATDNLIQKTI
                     REEFKSCTMLIIAHRLNTIIDSDKIMVLDSGRLKEYDEPYVLLQNKESLFYKMVQQTG
                     KANAEYLRSLVLGEKEVSKVDEVNSVSSERQKLDKNEAASSLIKKEEQEVPLSRLASL
                     NKPEIPVLLLGSIAAIINGVIFPIFGILISSAIKTFYEPPHELRKDSRFWALMFVVLG
                     VISFITIPARTYFFAVAGCKLIQRIRSMCFEKVVHMEVGWFDEPEHSSGAIGARLSAD
                     AASVRSLVGDALALIVQNISTAVTGLIIAFTANWRLALIVIAVLPLIIGSTMQMKMLS
                     GFAADAKAMYEEASQVANDAVGSIRTVASFCAEEKVMELYKKKCEGPMKTGIRQGLIS
                     GIGFGFSFFLIFSSYALALWYGAKLILDKGYTFRAVINVFMAILTGAMALGQASSFSP
                     DYAKAKISAAHIFMIIERQPLIDSYSEEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQ
                     GLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGMVSIDGQEIRSLNVQWLRAQ
                     MGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVAAAKEANIHAFIESLPNKYSTKVG
                     SKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKARVGRTTI
                     VIAHRLSTIRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAA
                     DESKSEIDALEMSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNLVLSIFSLFFIMVATL
                     YAMVKTTRGDGATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFYKQRDLLFYPAWAYAI
                     PSWILKIPITFIEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLLVNQMASALFRFIGAV
                     GRNMIVANTFGSFALLVIFALGGFILSRDDIKKWWIWGYWISPLMYAQNAISVNEFLG
                     HSWSHVVPNSTESLGVQVLKSRGFFTEAYWYWIGVGALIGFTLLFNILFTLALTYLNP
                     LGKPQAVISEEELKERDSRRGGEMQELKSGTSHGDSADAERAKMRLPESTSHVDTDIR
                     LLEMPHSEHERSIVQGQVETQDNKGPVQHVVKVSIWWMAYGVHIVIHSLILGFIFQFL
                     QIFSDWWLAYETPADKSKPFNPTVLIEVIYVIVQSFCLLKAMRAVVFVYVTLKASQNL
                     HNKMFESILRAPMSFFDSTPSGRILNRASTDQSAVDLDIPFQLGFFAFSAIQLLGIIA
                     VMSQVAWQVFIVFIPVIAISIWYQQYYISSARELSRLVGVCKAPVIQHFAESISGAMT
                     IRSFDQTTRFNAALMDHTQDASIFFHSFAMNEWLIQRMETLSAIVLASSLIMVLLPPG
                     TFNPGFVGMALSYGLSLNMVLFWAVYMACFIENKMVSVERIKQFTDIPSEAPWEIKDS
                     APPPSWPQVGRVEFRDYGLRYREDLDLVLKHINVDIKSGEKVGIVGRTGAGKSSMGVA
                     LFRLVEPAAGKIIIDGVDISTIGLHDLRSRLSIIPQDPTMFEGTVRSNLDPLEEHTDE
                     QIWEVLDKCQLGDEVRKKEGKLDSTVVENGENWSVGQRQLVCLGRVLLKRSKILVLDE
                     ATASVDTATDNLIQKTIREEFKSCTMLIIAHRLNTIIDSDKIMVLDSGRLKEYDEPYV
                     LLQNKESLFYKMVQQTGKANAEYLRSLVLGEKEVSKVDEVNSVSSERQKLDKNEAASS
                     LIKKEEQEVPLSRLASLNKPEIPVLLLGSIAAIINGVIFPIFGILISSAIKTFYEPPHE
                     LRKDSRFWALMFVVLGVISFITIPARTYFFAVAGCKLIQRIRSMCFEKVVHMEVGWFD
                     EPEHSSGAIGARLSADAASVRSLVGDALALIVQNISTAVTGLIIAFTANWRLALIVIAV
                     LPLIIGSTMQMKMLSGFAADAKAMYEEASQVANDAVGSIRTVASFCAEEKVMELYKKK
                     CEGPMKTGIRQGLISGIGFGFSFFLIFSSYALALWYGAKLILDKGYTFRAVINVFMAI
                     LTGAMALGQASSFSPDYAKAKISAAHIFMIIERQPLIDSYSEEGLMPNTLEGNVTFGE
                     VVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGMVSID
                     GQEIRSLNVQWLRAQMGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVAAAKEANIH
                     AFIESLPNKYSTKVGSKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKV
                     VQEALDKARVGRTTIVIAHRLSTIRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLV
                     TMQTAGNEVELENAADESKSEIDALEMSSNDYRSSLIRAIYRSLGRSFVYVFRGTSNL
                     VLSIFSLFFIMVATLYAMVKTTRGDGATVMGSLYAVTMIIVSGYVEVSMTLAKLPVFY
                     KQRDLLFYPAWAYAIPSWILKIPITFIEVAIWVFLTYYVIGFDPNVGRLFKQYLLLLL
                     VNQMASALFRFIGAVGRNMIVANTFGSFALLVIFALGGFILSRDDIKKWWIWGYWISP
                     LMYAQNAISVNEFLGHSWSHVVPNSTESLGVQVLKSRGFFTEAYWYWIGVGALIGFTL
                     LFNILFTLALTYLNPLGKPQAVISEEELKERDSRRGGEMQELKSGTSHGDSADAERAK
                     MRLPESTSHVDTDIRLLEMPHSEHERSIVQGQVETQDNKGPVQHVVKVSIWWMAYGVH
                     IVIHSLILGFIFQFLQIFSDWWLAYETPADKSKPFNPTVLIEVIYVIVQSFCLLKAMR
                     AVVFVYVTLKASQNLHNKMFESILRAPMSFFDSTPSGRILNRASTDQSAVDLDIPFQL
                     GFFAFSAIQLLGIIAVMSQVAWQVFIVFIPVIAISIWYQQYYISSARELSRLVGVCKA
                     PVIQHFAESISGAMTIRSFDQTTRFNAALMDHTQDASIFFHSFAMNEWLIQRMETLSA
                     IVLASSLIMVLLPPGTFNPGFVGMALSYGLSLNMVLFWAVYMACFIENKMVSVERIKQ
                     FTDIPSEAPWEIKDSAPPPSWPQVGRVEFRDYGLRYREDLDLVLKHINVDIKSGEKVG
                     IVGRTGAGKSSMGVALFRLVEPAAGKIIIDGVDISTIGLHDLRSRLSIIPQDPTMFEG
                     TVRSNLDPLEEHTDEQIWEVLDKCQLGDEVRKKEGKLDSTVVENGENWSVGQRQLVC
                     LGRVLLKRSKILVLDEATASVDTATDNLIQKTIREEFKSCTMLIIAHRLNTIIDSDKI
                     MVLDSGRLKEYDEPYVLLQNKESLFYKMVQQTGKANAEYLRSLVLGEKEVSKVDEVN
                     SVSSERQKLDKNEAASSLIKKEEQEVPLSRLASLNKPEIPVLLLGSIAAIINGVIFPI
                     FGILISSAIKTFYEPPHELRKDSRFWALMFVVLGVISFITIPARTYFFAVAGCKLIQR
                     IRSMCFEKVVHMEVGWFDEPEHSSGAIGARLSADAASVRSLVGDALALIVQNISTAVT
                     GLIIAFTANWRLALIVIAVLPLIIGSTMQMKMLSGFAADAKAMYEEASQVANDAVGSI
                     RTVASFCAEEKVMELYKKKCEGPMKTGIRQGLISGIGFGFSFFLIFSSYALALWYGAK
                     LILDKGYTFRAVINVFMAILTGAMALGQASSFSPDYAKAKISAAHIFMIIERQPLIDS
                     YSEEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKST
                     VVQLLERFYDPLAGMVSIDGQEIRSLNVQWLRAQMGIVSQEPILFDCSIAENIAYGDN
                     SRVVSQEEIVAAAKEANIHAFIESLPNKYSTKVGSKGTQLSGGQKQRIAIARALVRQP
                     HILLLDEATSALDTESEKVVQEALDKARVGRTTIVIAHRLSTIRNADVIAGFDDGVIV
                     EKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDYRSSLIRA
                     IYRSLGRSFVYVFRGTSNLVLSIFSLFFIMVATLYAMVKTTRGDGATVMGSLYAVTMI
                     IVSGYVEVSMTLAKLPVFYKQRDLLFYPAWAYAIPSWILKIPITFIEVAIWVFLTYYV
                     IGFDPNVGRLFKQYLLLLLVNQMASALFRFIGAVGRNMIVANTFGSFALLVIFALGGF
                     ILSRDDIKKWWIWGYWISPLMYAQNAISVNEFLGHSWSHVVPNSTESLGVQVLKSRGF
                     FTEAYWYWIGVGALIGFTLLFNILFTLALTYLNPLGKPQAVISEEELKERDSRRGGEM
                     QELKSGTSHGDSADAERAKMRLPESTSHVDTDIRLLEMPHSEHERSIVQGQVETQDNK
                     GPVQHVVKVSIWWMAYGVHIVIHSLILGFIFQFLQIFSDWWLAYETPADKSKPFNPTV
                     LIEVIYVIVQSFCLLKAMRAVVFVYVTLKASQNLHNKMFESILRAPMSFFDSTPSGRI
                     LNRASTDQSAVDLDIPFQLGFFAFSAIQLLGIIAVMSQVAWQVFIVFIPVIAISIWYQ
                     QYYISSARELSRLVGVCKAPVIQHFAESISGAMTIRSFDQTTRFNAALMDHTQDASIF
                     FHSFAMNEWLIQRMETLSAIVLASSLIMVLL
```

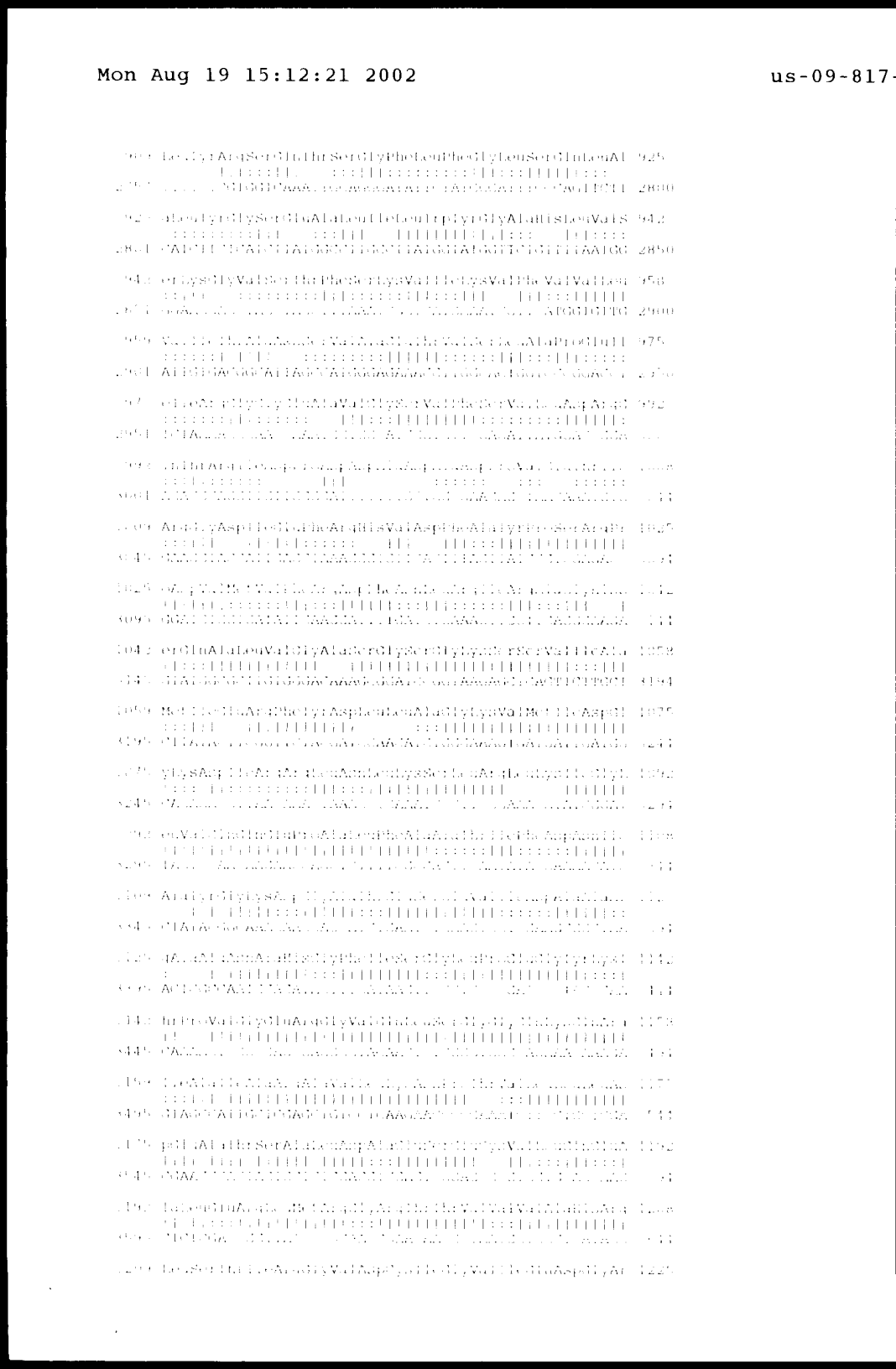






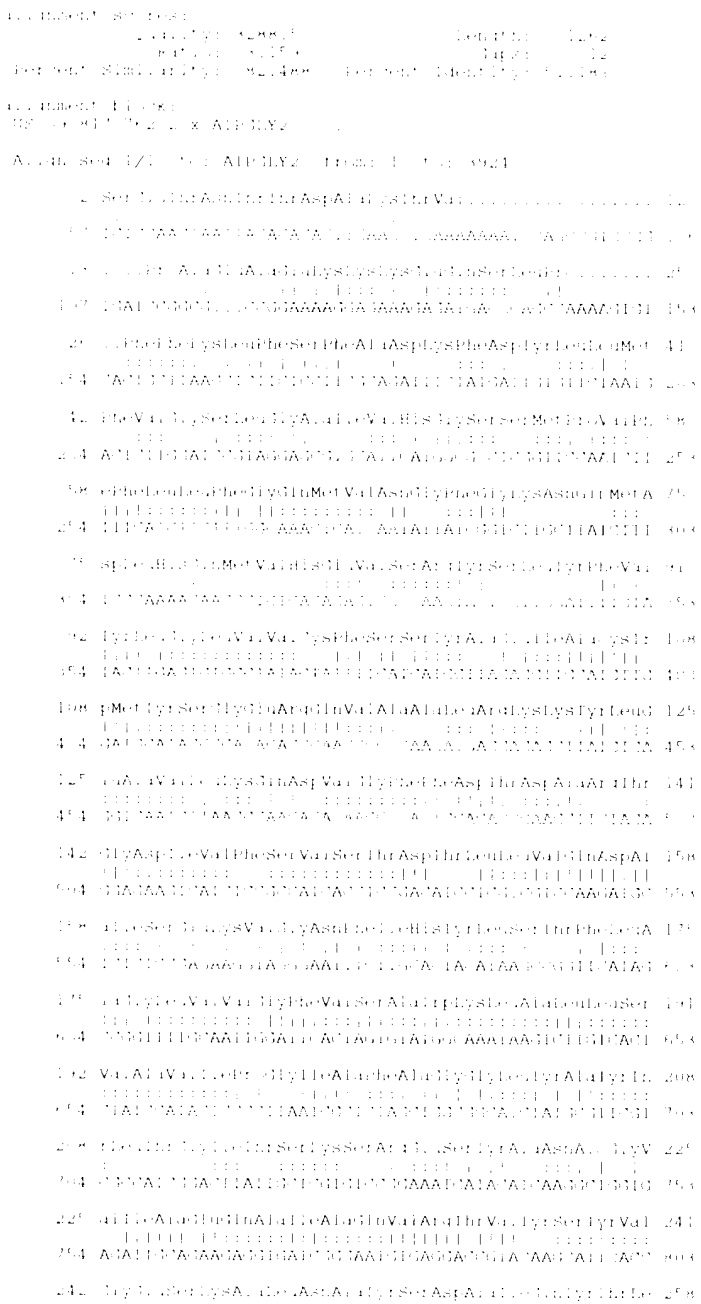


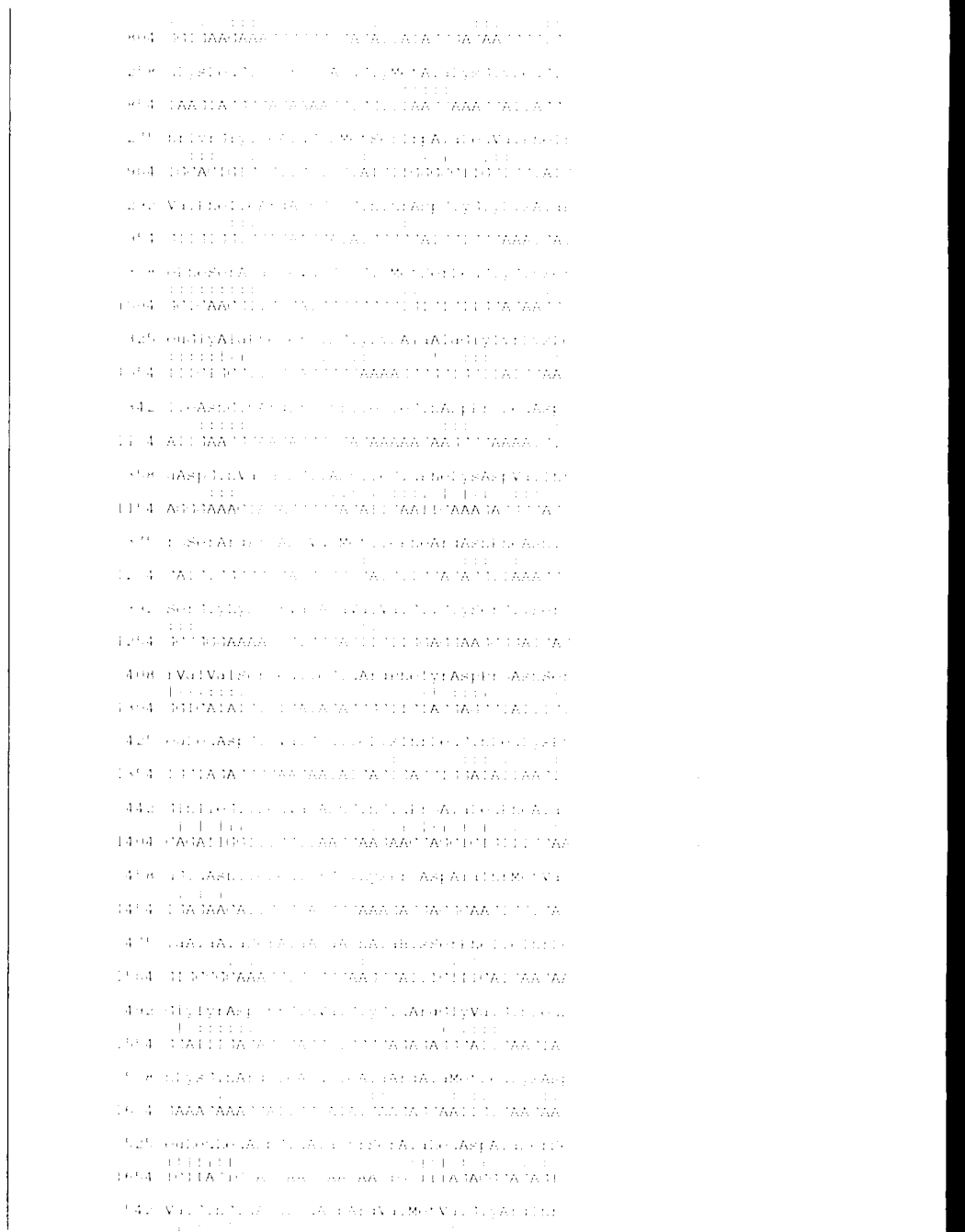


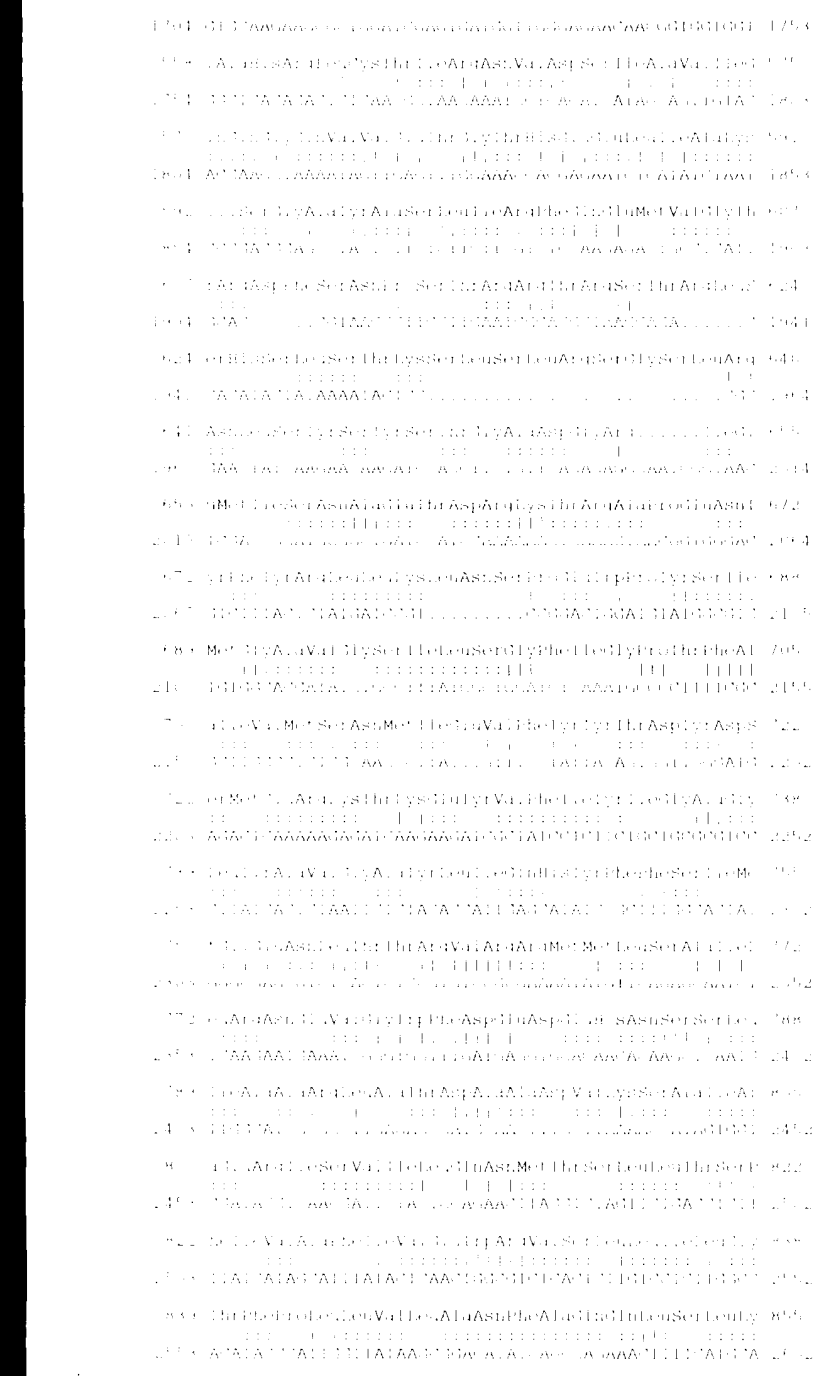


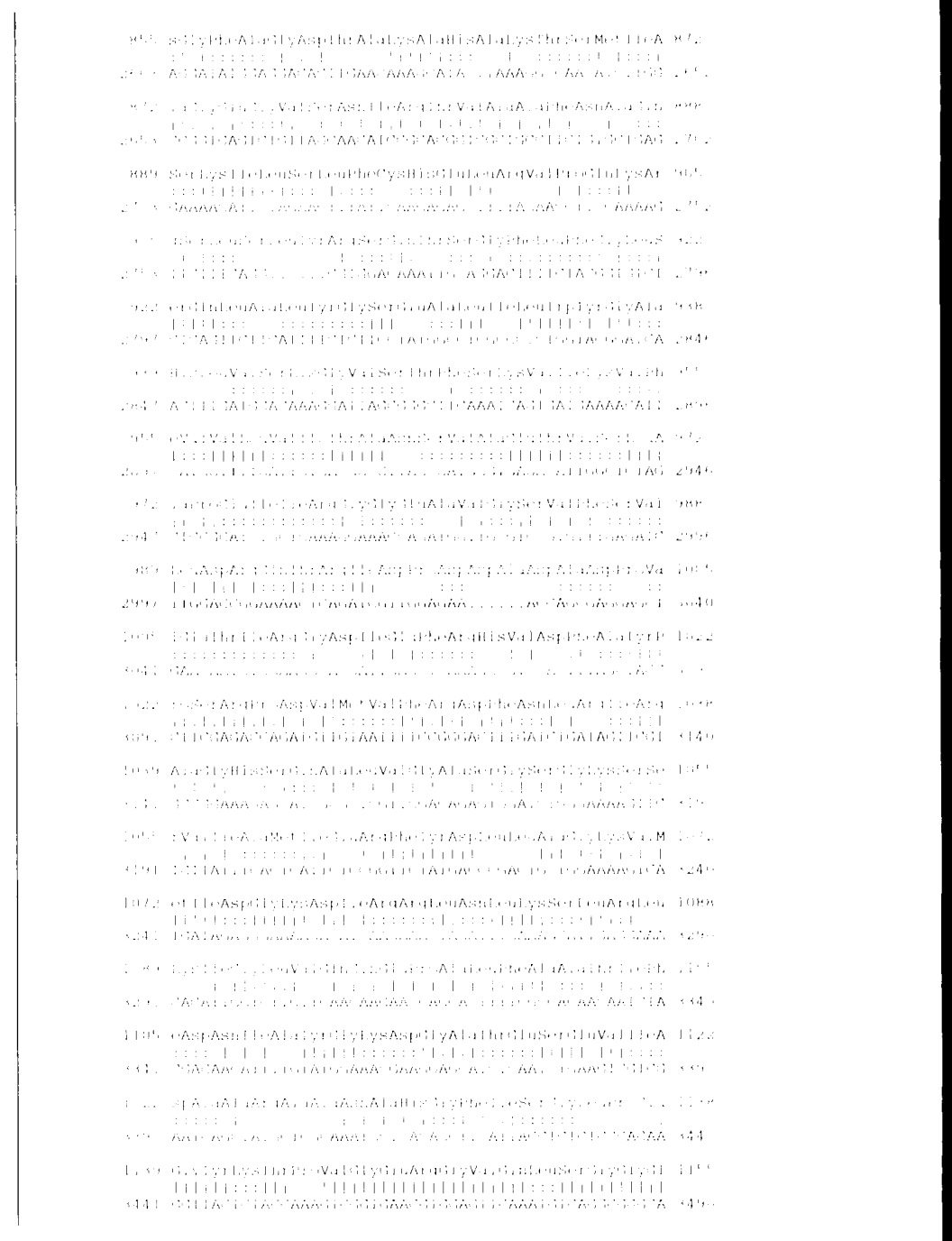
```
Seq_name: qb_pl:AI19LY2

seq_documentation_block:
LOCUS       AI19LY2                 3924 bp    mRNA    linear   PLN 12-MAY-1997
DEFINITION  A.thaliana mRNA for Polycoprotein 2.
ACCESSION   Y19228
VERSION     Y19228.1  GI:2108252
KEYWORDS    Polycoprotein-2; pgp2 gene.
SOURCE      Chate cross.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE   1  (bases 1 to 3924)
  AUTHORS   Marques,J.P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3924)
  AUTHORS   Marques,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-1997) Jose Pedro Marques, Max-Planck Institut Fuer
            Zuechtungsforsc, Genetic Principles of Plant Breeding,
            Carl Von Linne Weg 10, 50829 Koeln, FRG
FEATURES             Location/Qualifiers
     source          1..3924
                     /organism "Arabidopsis thaliana"
                     /db_xref "taxon:3702"
                     /chromosome "4"
                     /map "72.4 cM"
                     /cell_line "Columbia"
     gene            1..3794
                     /gene "pgp2"
     CDS             1..3794
                     /gene "pgp2"
                     /codon_start=1
                     /product "Polycoprotein 2"
                     /protein_id="CAA21277.1"
                     /db_xref="GI:2108252"
                     /translation="MSPCSTAPSPEPIMSIPPVSLLPLPLAPLVSVLMLSCVSA
                     CWQPSRSRAYSAMLSQSLEELAGTGVSAFSVGAGSSSSSVGLGEVG
                     NPLGFLGGTASFAFSFLVWSLSVLLHVLLALAGIVAFVAVGLLAVWKSYLK
                     AGLIASVVSCVLVVAFCLSAVELTLAEMSTGFPFRAGSGFFLSVFLLGLW
                     FFLLSGLVRLGAGRAVLVAFLLMSVGSFAVLALRVVSLFGDGV
                     TTIKENLVKGLATAEETKARLSTAIGCTRRELEIELVGERGTQLSGQQKQKI
                     LAVVHEGKIVLFGNPENLLNKGAYCSLLKSETASDQSNPGLNKTLSEFHSIKYSK
                     QMSLEAAVCLSAVSTYSMGEDLRELEKEAIDA"
BASE COUNT      1064 a     860 c     948 g    1052 t
ORIGIN
```









[illegible]

Seq name: gb|pat|AX14[illegible]

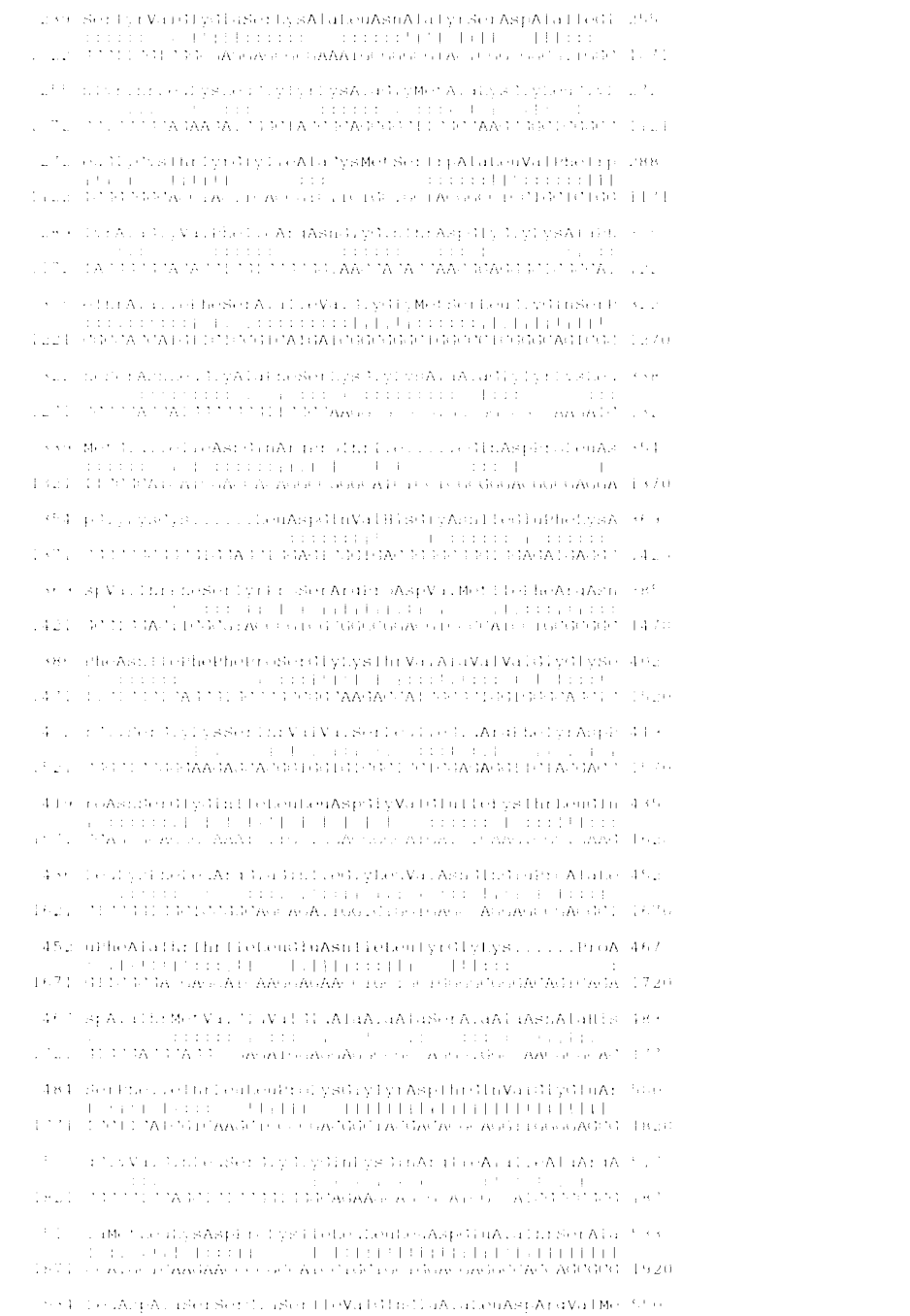
Seq documentation block:

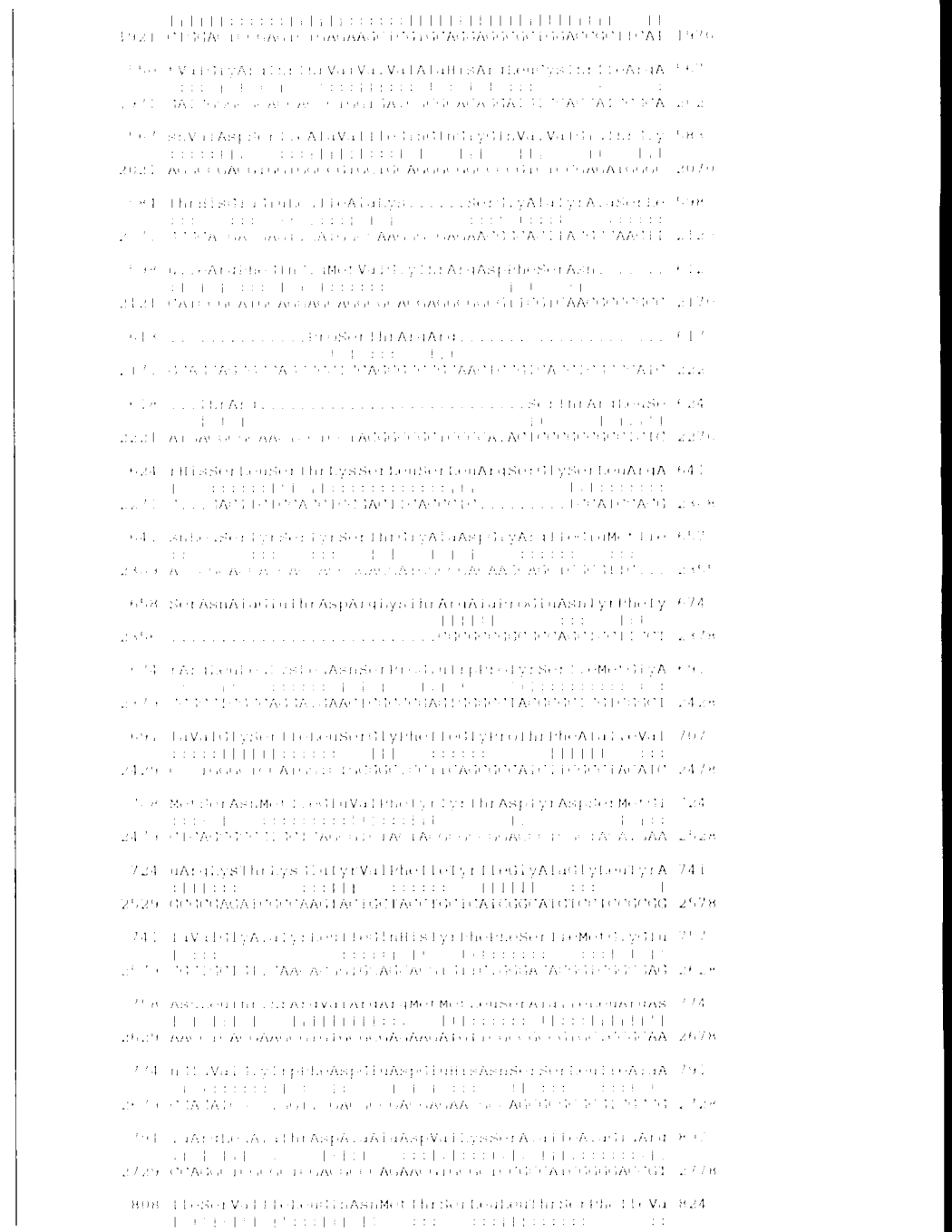
```
LOCUS       AX14[illegible]            4213 bp    DNA     linear   PAT 31-MAY-2001
DEFINITION  Sequence 8 from Patent WO0134818.
ACCESSION   AX14[illegible]
VERSION     AX14[illegible]  GI:1428[illegible]
KEYWORDS    .
SOURCE      Sorghum [illegible]
  ORGANISM  Sorghum [illegible]
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum
REFERENCE   1  (bases 1 to 4213)
  AUTHORS   Johal,G.S. and Multani,D.S.
  TITLE     Sorghum dwarfing genes and methods of use
  JOURNAL   Patent: WO 0134818-A 8 17-MAY-2001;
            Johal, Gurmukh S. (US) ; Multani, Dilbag S. (US)
FEATURES             Location/Qualifiers
     source          1..4213
                     /organism="Sorghum [illegible]"
                     /db_xref="taxon:4[illegible]"
     CDS             1..42[illegible]
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC[illegible]"
                     /db_xref="GI:1428[illegible]"
                     /translation="[illegible]"
```

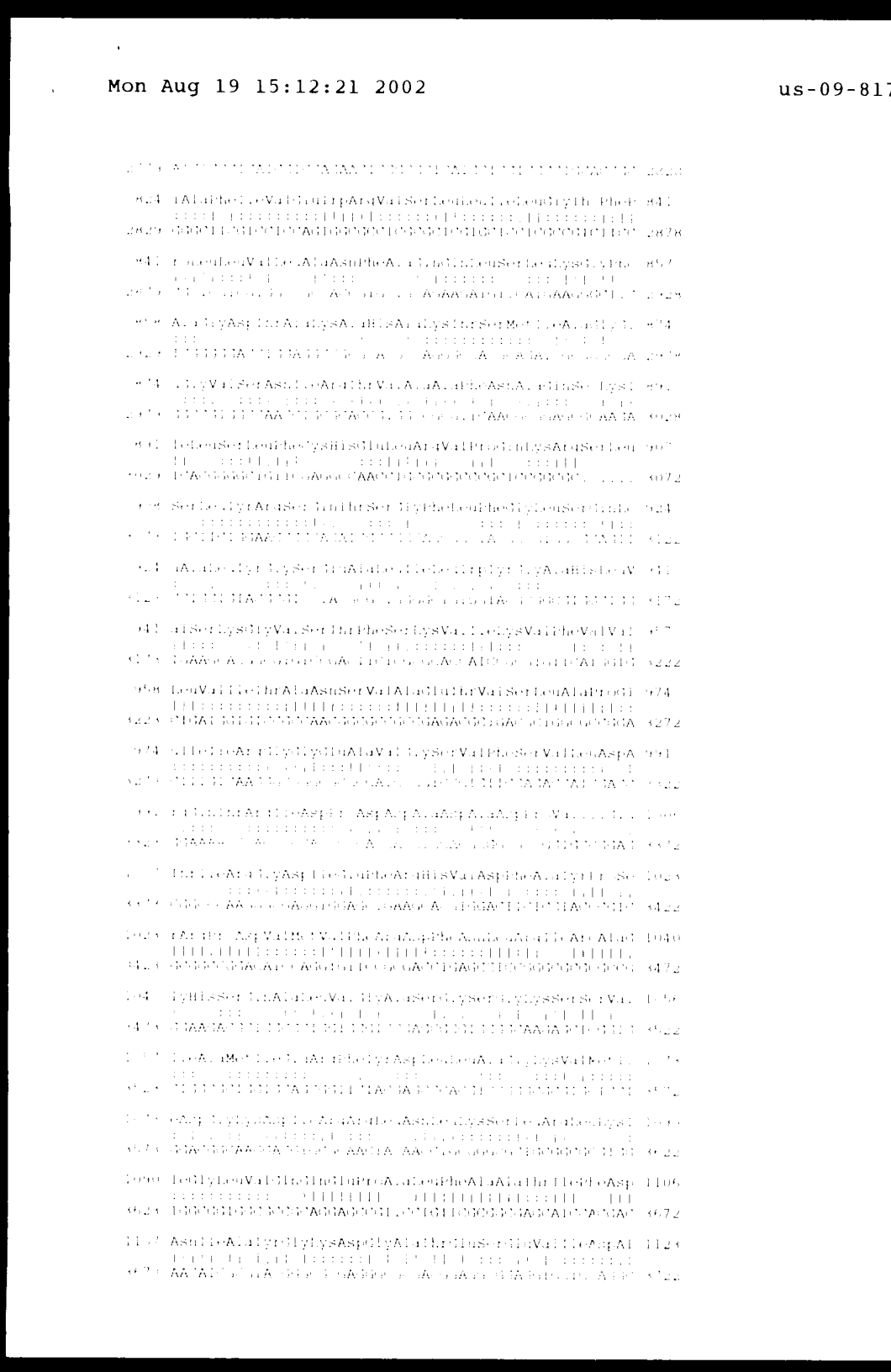
BASE COUNT [illegible]

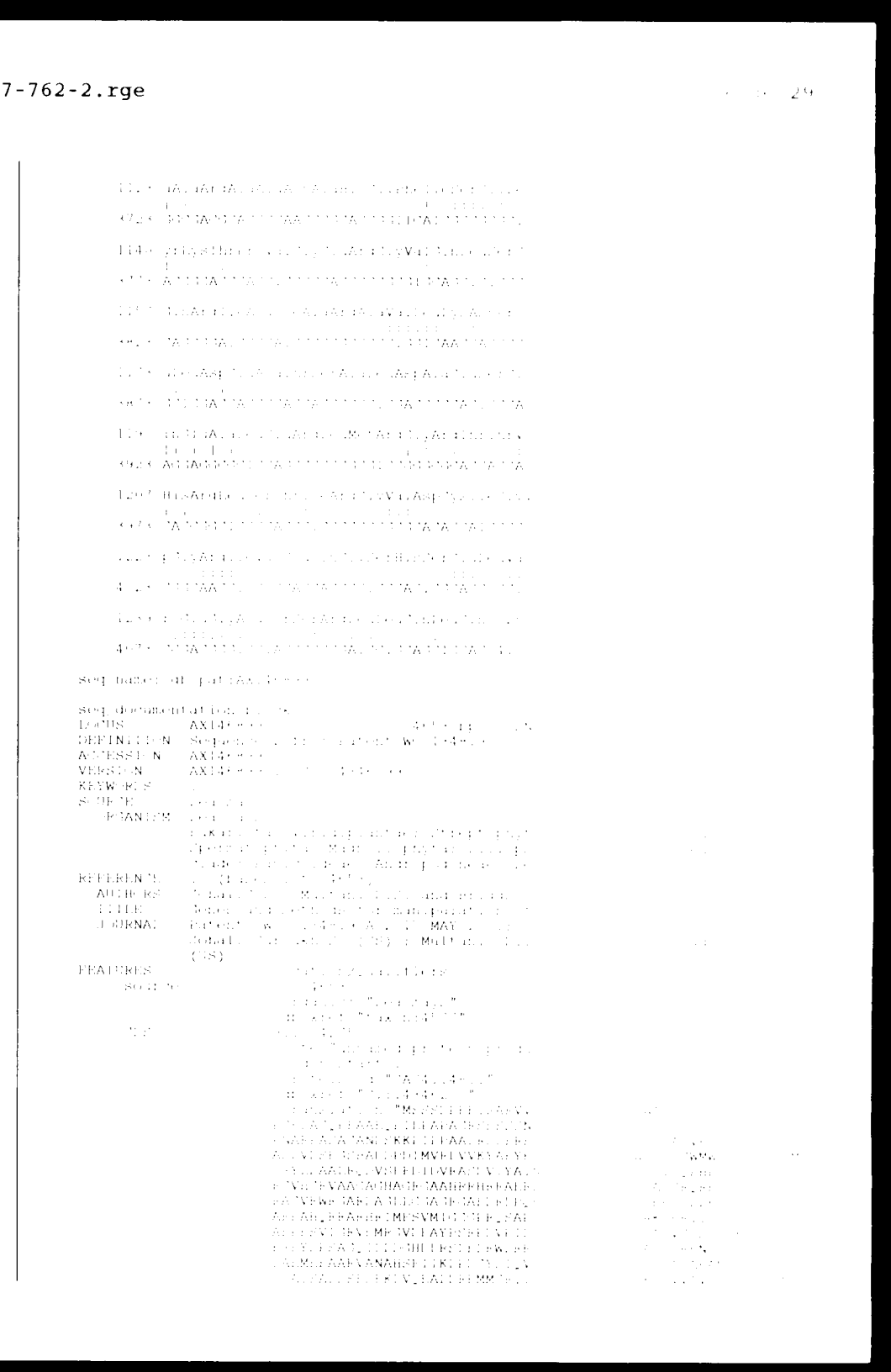
ORIGIN

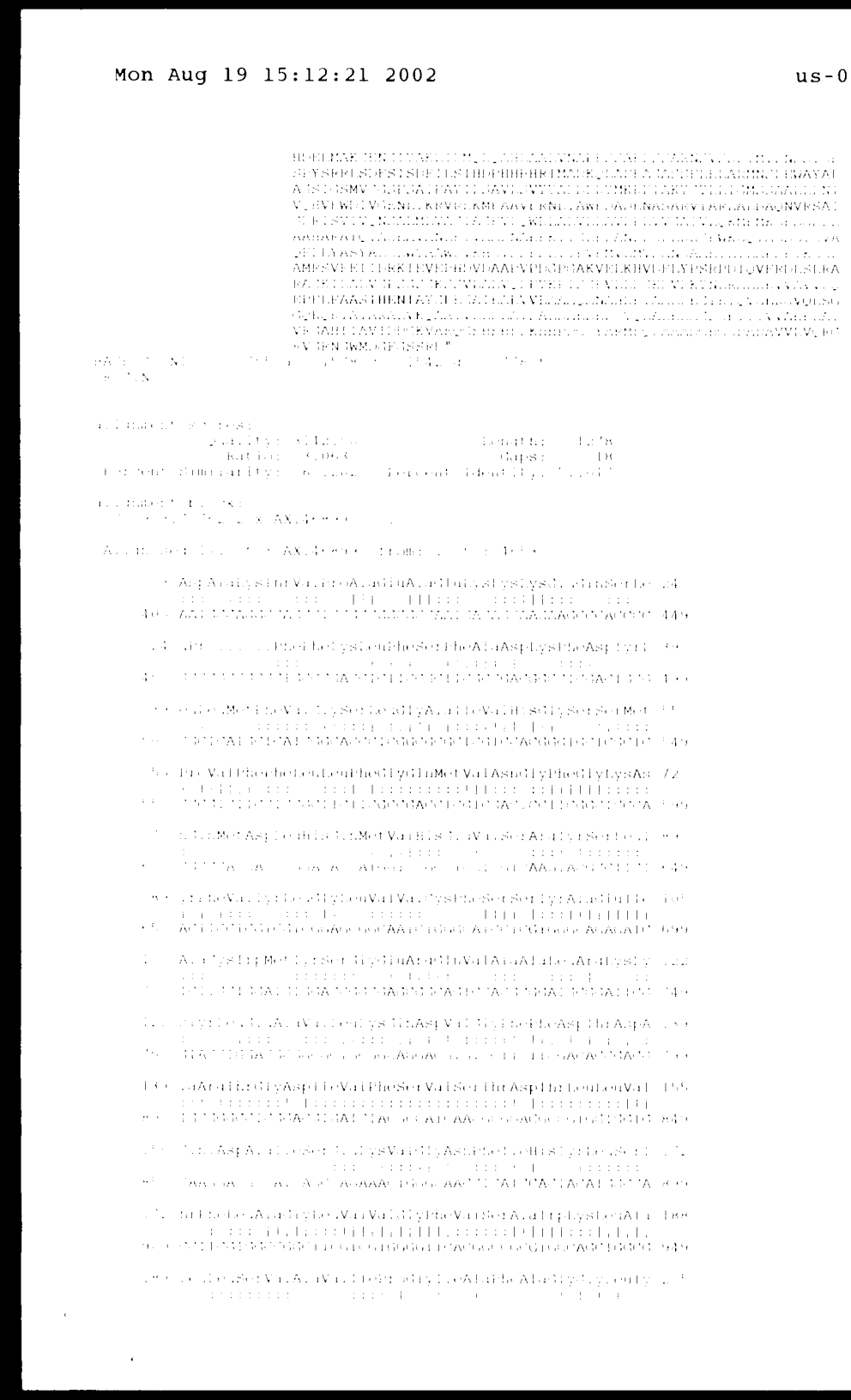
[illegible]

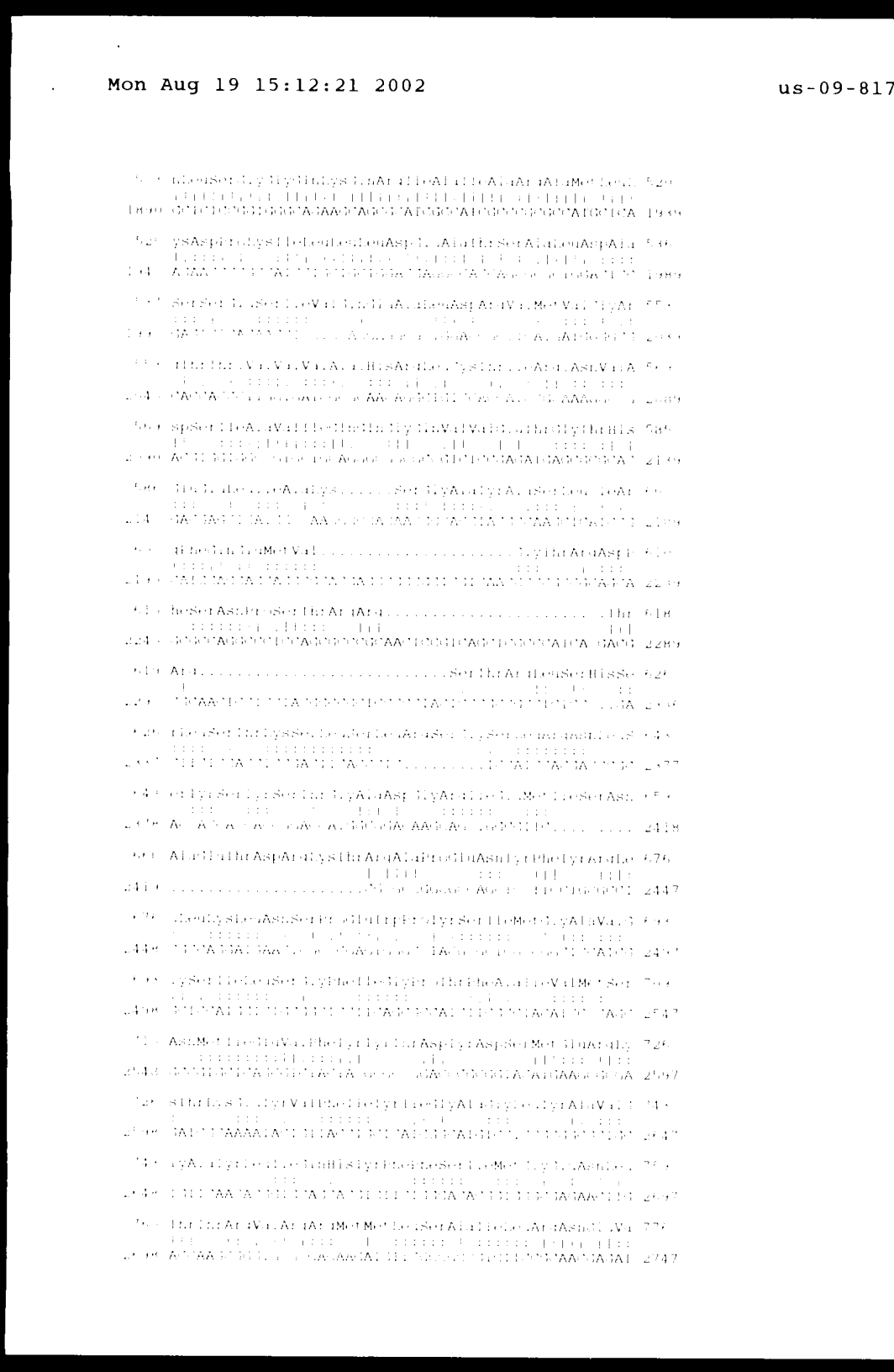


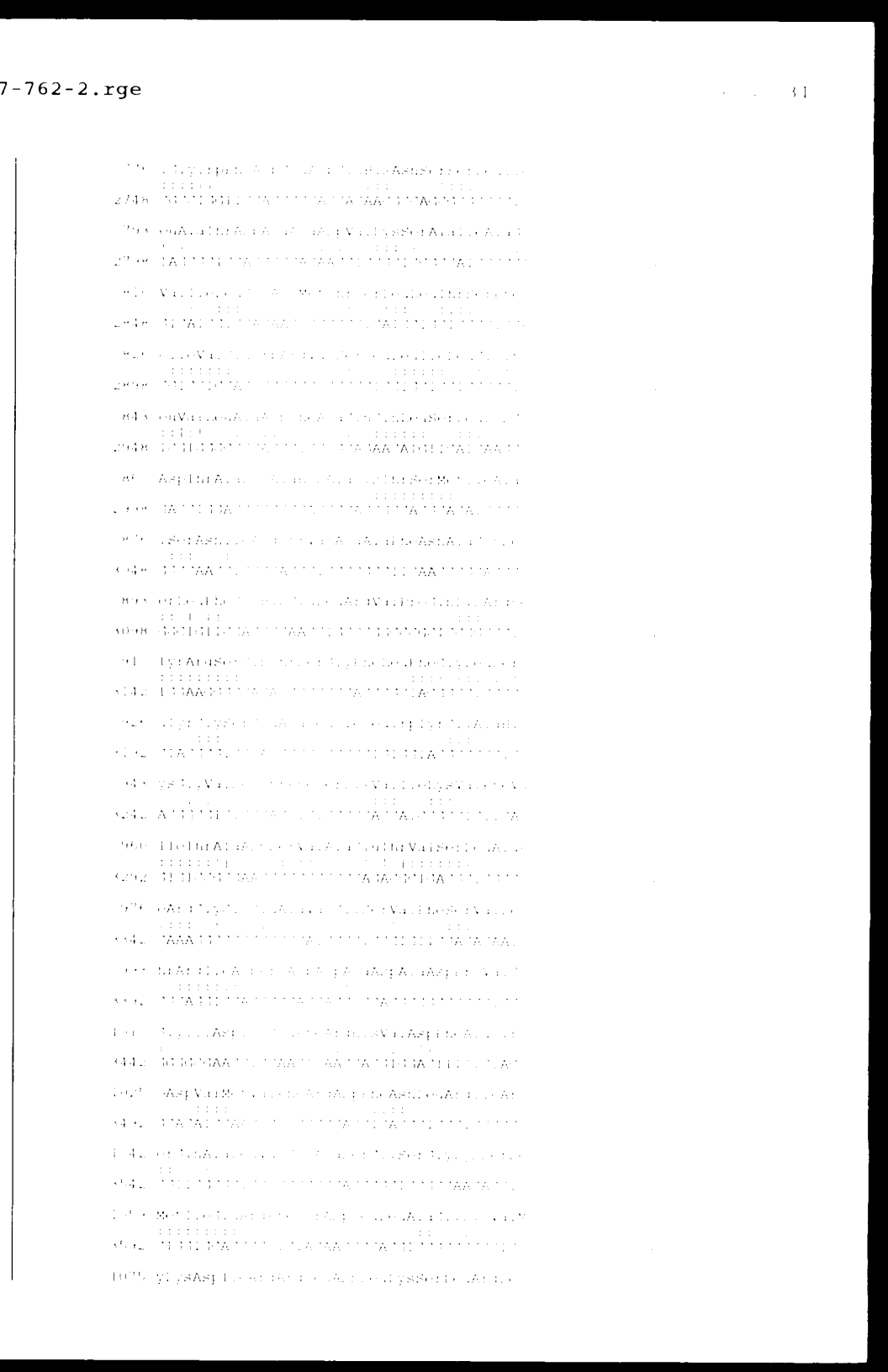






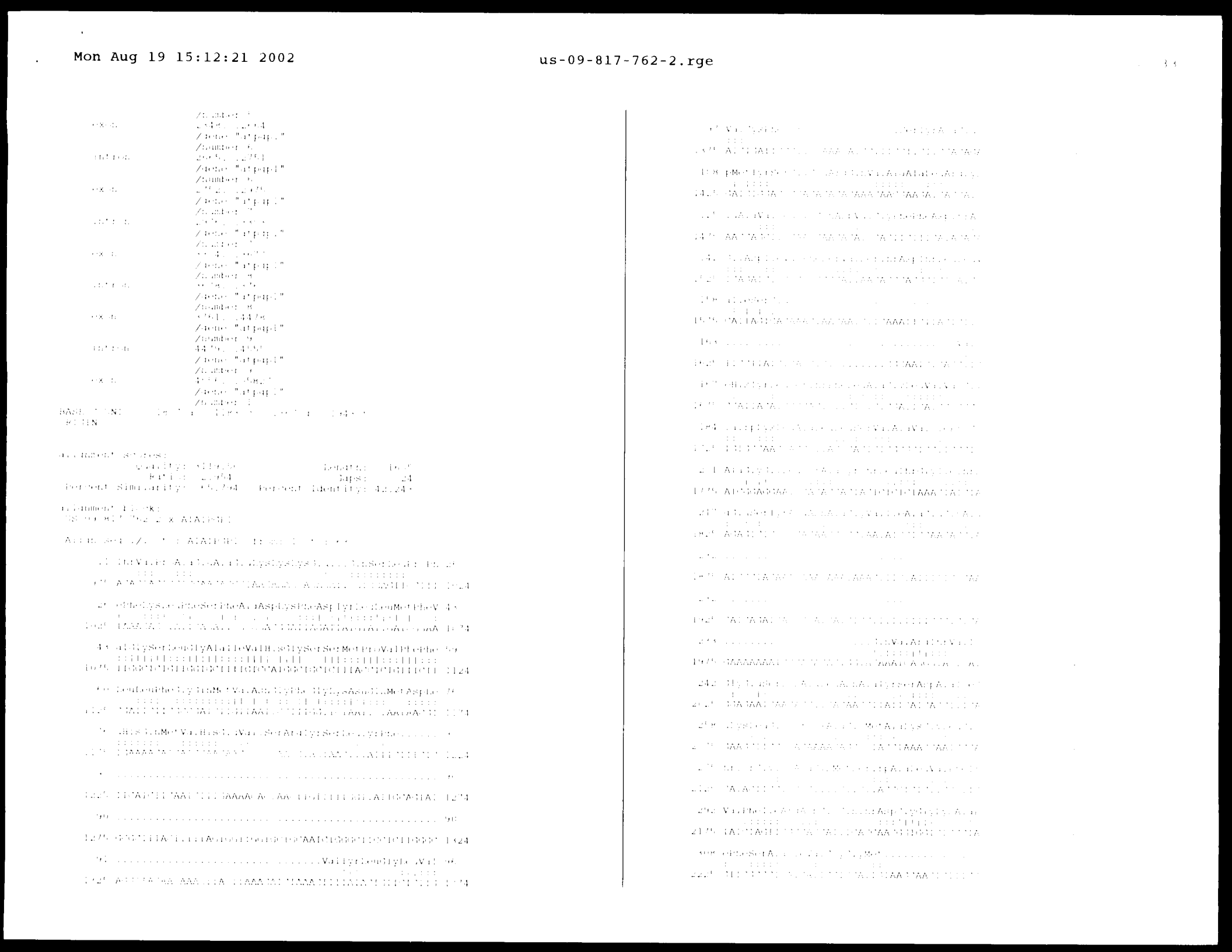


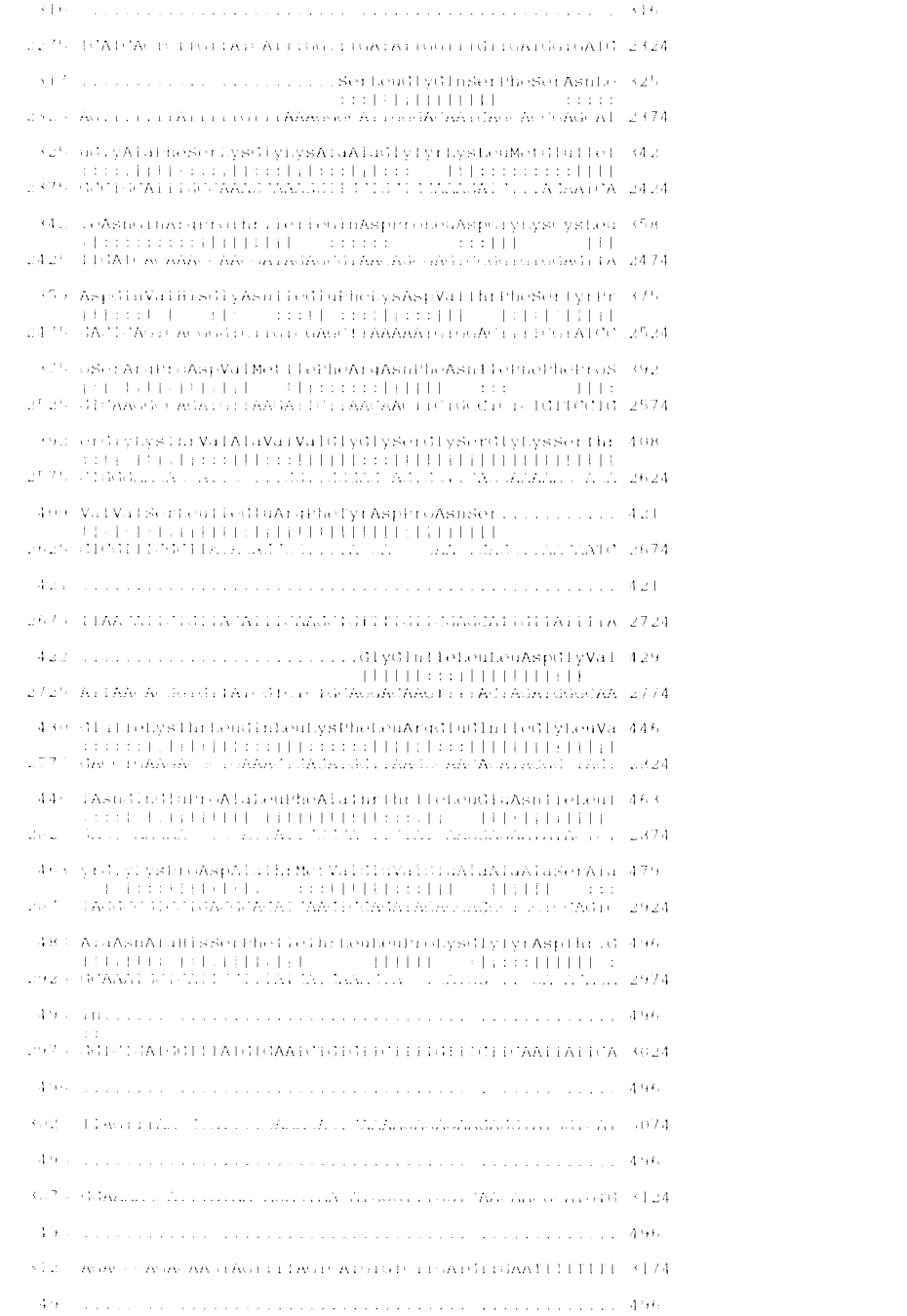


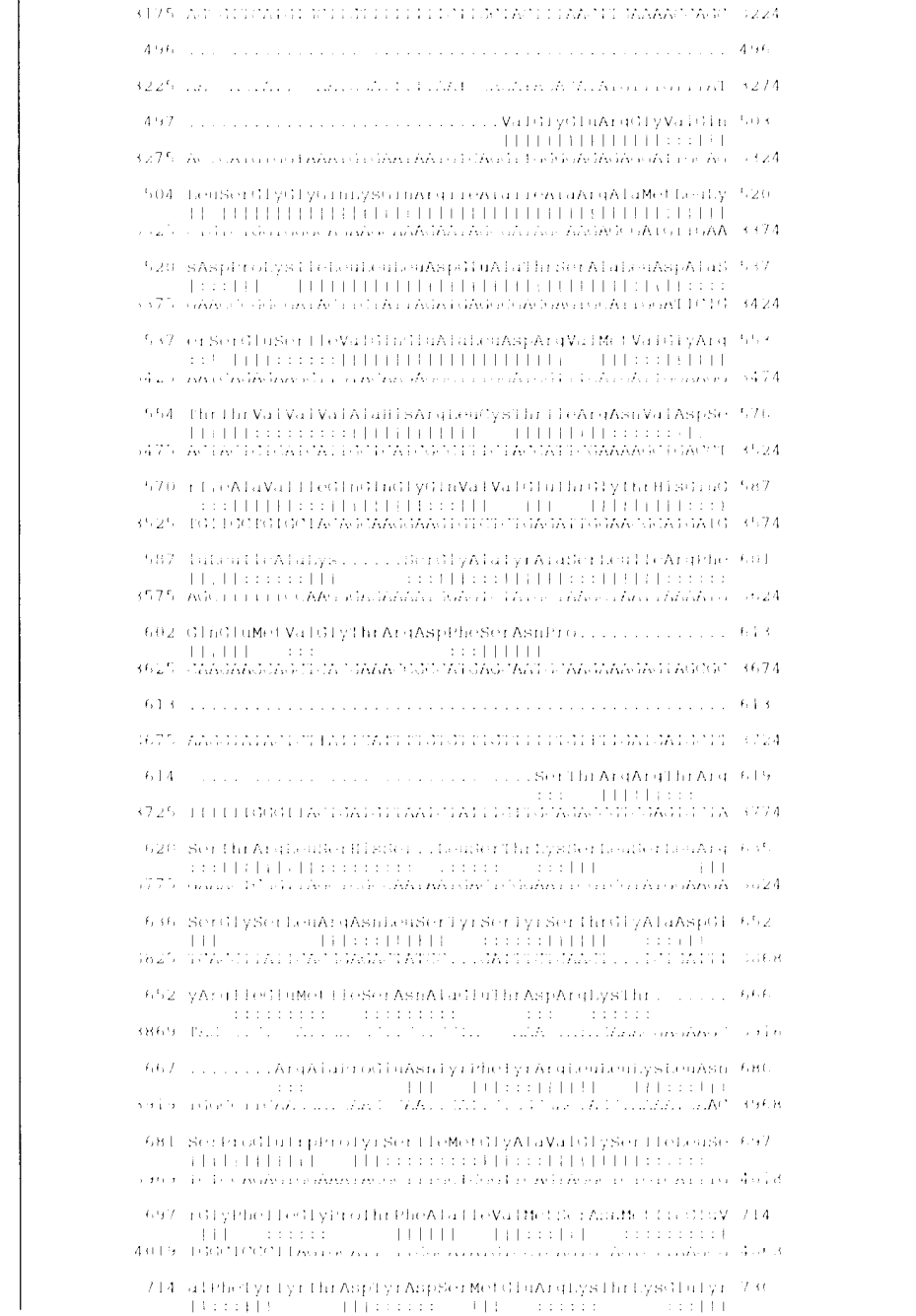


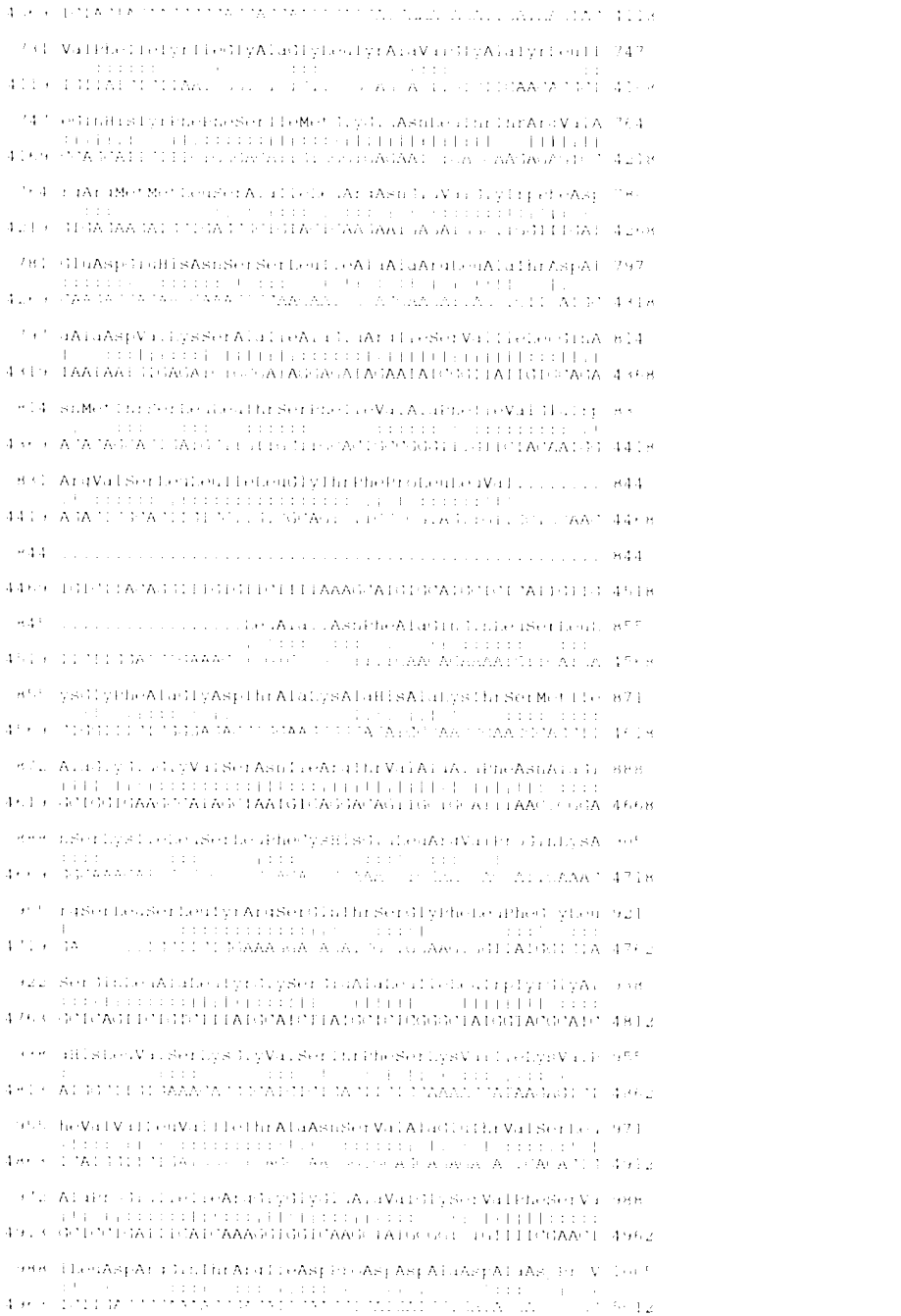
```
set name: gb_pl:AtATPGPI

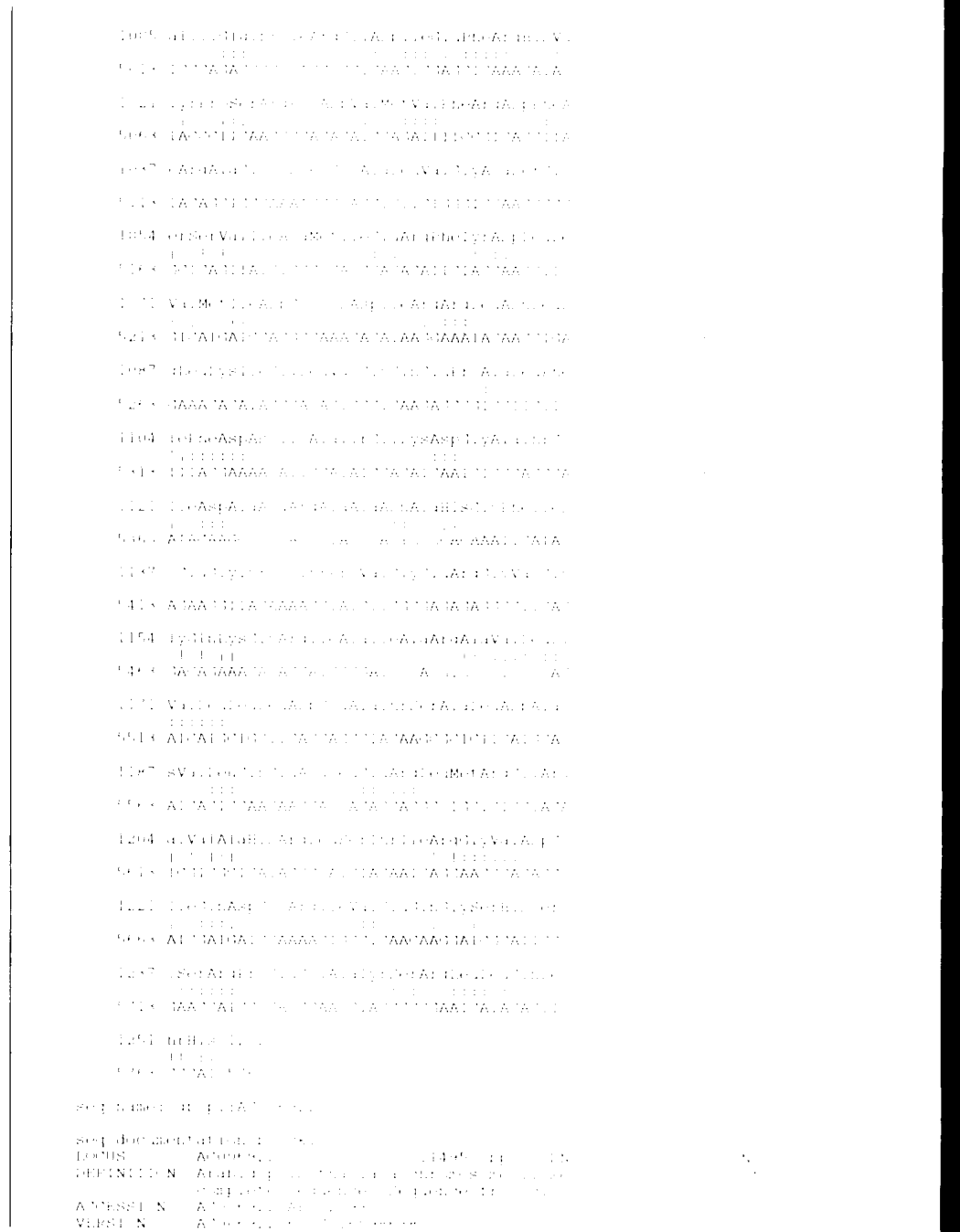
set documentation block.
LOCUS       AtATPGPI                6300 bp    DNA     linear   PLN 09-NOV-1998
DEFINITION  A.thaliana atppl gene for Pglycoprotein, homologous to mammalian
            mdr genes.
ACCESSION   X61378
VERSION     X61378.1  GI:16185
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 6300)
  AUTHORS   Sudler,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15 AUG 1991) K. Sudler, Inst of Plant Biology,
            University of Zuerich, Zollikerstrasse 107, CH-8008 Zuerich,
            SWITZERLAND
REFERENCE   2  (bases 1 to 6300)
  AUTHORS   Sudler,K. and Hertig,C.
  TITLE     Structure of an mdr like gene from Arabidopsis thaliana:
            Evolutionary implications
  JOURNAL   J. Biol. Chem. 267 (9), 5982-5988 (1992)
  MEDLINE   92210050
FEATURES             Location/Qualifiers
     source          1..6300
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="lambda EMBL3"
     TATA_signal     759..765
                     /gene="atppl"
     gene            759..5827
                     /gene="atppl"
     exon            794..1202
                     /gene="atppl"
                     /number=1
     mRNA            join(794..1202,1272..1326,1412..1587,1661..1861,
                     1989..2248,2348..2664,2752..2975,3304..3677,3764..4478,
                     4556..5827)
                     /gene="atppl"
                     /evidence=experimental
     CDS             join(939..1202,1272..1326,1412..1587,1661..1861,
                     1989..2248,2348..2664,2752..2975,3304..3677,3764..4478,
                     4556..5827)
                     /gene="atppl"
                     /note="homologous to mammalian mdr gene,contains
                     ATP binding cassette"
                     /codon_start=1
                     /product="P-glycoprotein"
                     /protein_id="CAA43646.1"
                     /db_xref="GI:849833"
                     /translation="MEDEASAFGSKNVRRGMDLPKEDTIRGVVDKGNNPKLG
                     LVKAWWFSFMFMLCKRSLKNTINQDSFYSLGLAMQSVLVQSISNLMLNKLGQS
                     SFATIDSGRIINRFSKDIGHLDDLLPLTFLDFIQTLLQVVGVVSVISVVAPWV
                     LIAVVPLLVAFIMVRAYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLWTLRAF
                     GRQPYFETLFHKALNLHTANWFLYLSTLRWFQMRIEMIFVIFFIAVTFISILT
                     TGEGEGRVGIILTLAMNIMSTLQWAVNSSIDVDSLMRSVSRVFKFIDMPTEG
                     KPTKSTKPYKNGQLSKVMIIENSHVKKDDIWPSGGQMTVKDLTAKYTEGGNA
                     ILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRLLNTEGEIQIDGVSWDSIT
                     LQQWRKAFGVIPQKVFIFSGTFRKNLDPYEQWSDQEIWKVADEVGLRSVIEQ
                     FPGKLDFVLVDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQI
                     IRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSL
                     FRQAISPSDRVKLFPHRNSSKCKSKPQIAALKEETEEEVQDTRL"
     intron          1203..1271
                     /gene="atppl"
                     /number=1
     exon            1272..1326
                     /gene="atppl"
                     /number=2
     intron          1327..1411
                     /gene="atppl"
                     /number=2
     exon            1412..1587
                     /gene="atppl"
                     /number=3
     intron          1588..1660
                     /gene="atppl"
                     /number=3
     exon            1661..1861
                     /gene="atppl"
                     /number=4
     intron          1862..1988
                     /gene="atppl"
                     /number=4
     exon            1989..2248
                     /gene="atppl"
                     /number=5
     intron          2249..2347
                     /gene="atppl"
```











```
KEYWORDS        .
SEGMENT         .
SOURCE          Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicots; rosids;
                Brassicales; Brassicaceae; Arabidopsis.
REFERENCE       1  (bases 1 to 119465)
  AUTHORS       Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.I., Town,C.D.,
                Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
                Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
                Koo,H.L., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
                Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
                Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
                Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
                Venter,J.C.
  TITLE         Sequence and analysis of chromosome 2 of the plant Arabidopsis
                thaliana
  JOURNAL       Nature 402 (6763), 761-768 (1999)
    PUBMED      10617198
  REMARK        .
  REFERENCE     2  (bases 1 to 119465)
  AUTHORS       .
  TITLE         Direct Submission
  JOURNAL       Submitted (20-MAR-2000) The Institute for Genomic Research, 9712
                Medical Center Dr., Rockville, MD 20850, USA
  REMARK        Gene names for this sequence version have been added at the
                request of the Arabidopsis Genome Initiative. The gene names
                for the nucleotide and annotation of chromosome 2 were derived from these
                 the individual clones on this chromosome after removing
                overlaps. For detailed information, please see the TIGR web site
                (http://www.tigr.org/tdb/at/at.html).

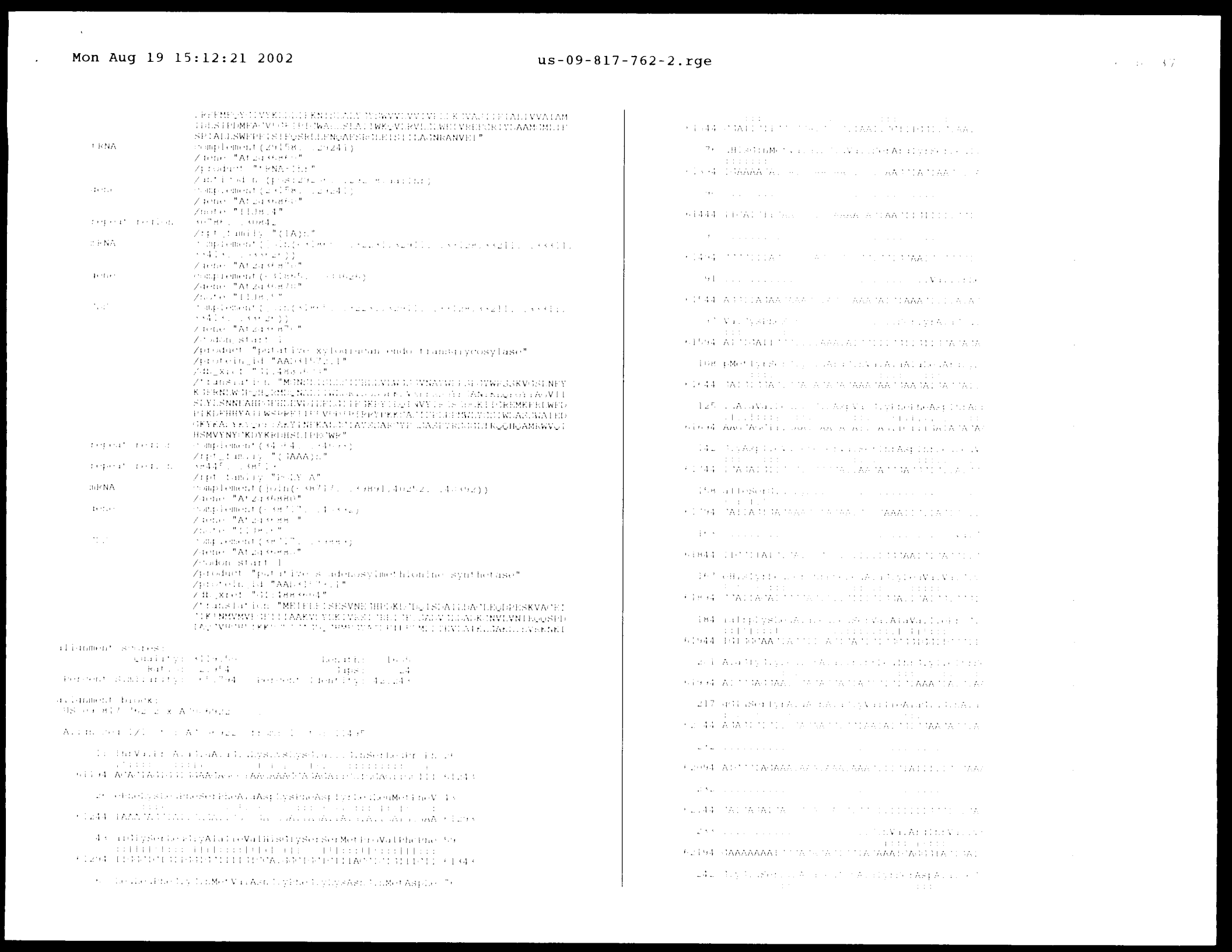
                Sixty were identified by a combination of three methods: Some
                proteins in Arabidopsis
                (http://arabidopsis.stanford.edu/proteins.html), the Arabidopsis
                University of Washington), Syngenta (Paris Europe,
                http://genome.cse.ucdavis.edu/ESTAnnotation.html), and MPIZuchtung
                (http://www.mpiz-koeln.mpg.de/). We used BLASTX to compare
                 sequence against a peptide database and plant EST
                databases at TIGR and manual curations based on these analyses.
                Annotated genes are named to indicate the level of confidence for
                their annotation. Genes with similarity to other proteins are named
                after the database hits. Genes without significant peptide
                similarity but with EST similarity are named as "expressed" proteins.
                Genes without protein or EST similarity, that are predicted by two
                 or more programs are most of their length are
                designated as "hypothetical" proteins. Genes encoding tRNAs are
                predicted by tRNAscan-SE (Sean Eddy,
                http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
                identified by RepeatMasker (Arian Smit,
                http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
                numbered in the order in the chromosome.

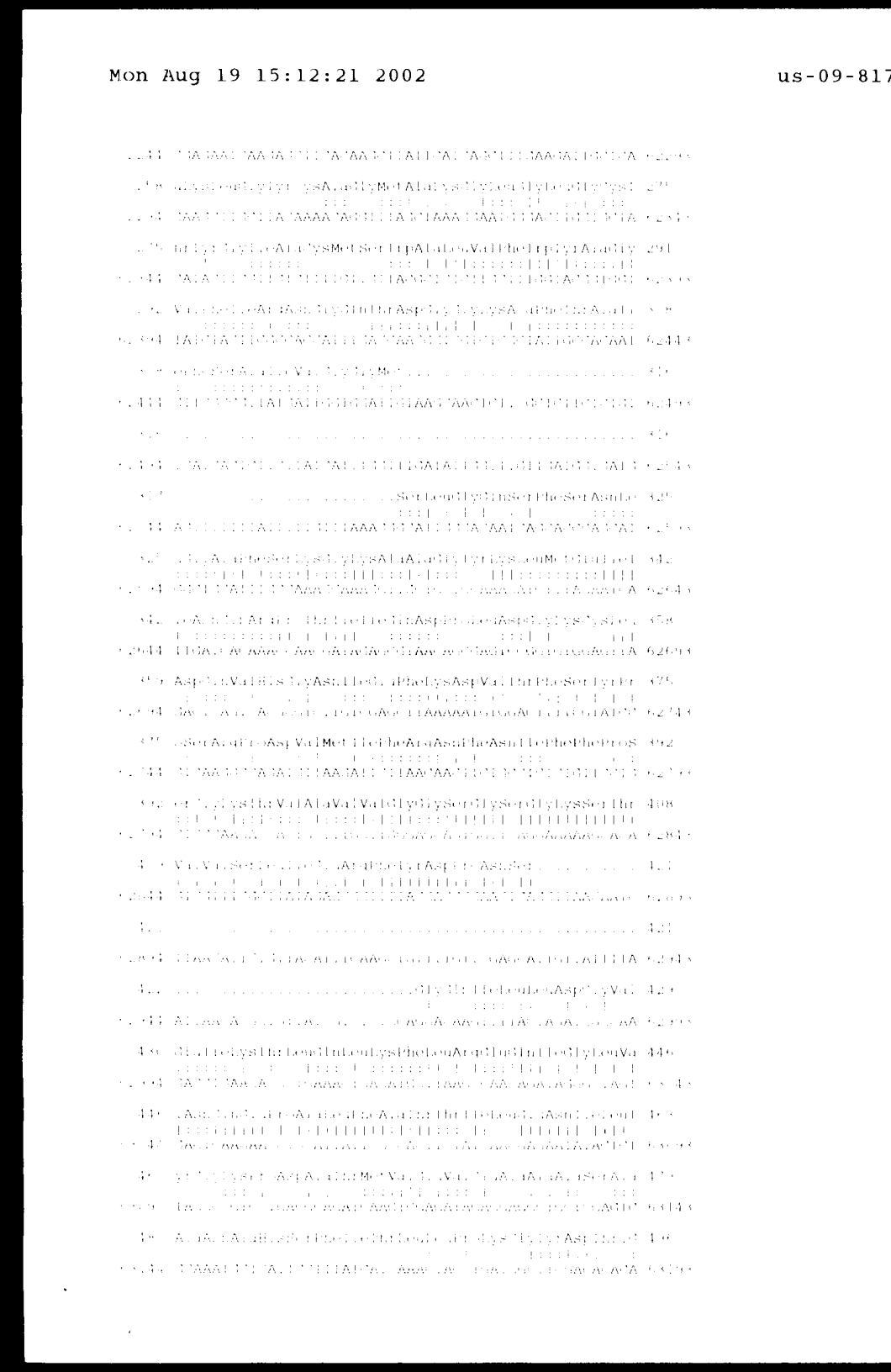
                We thank the DOE/Joint Genome Institute for sequencing clones
                F23A5, F5J6, T12A5, and T13G10, the EMSA group for sequencing clones
                F15B4, and Scott Jackson, Zhiqing Jiang, Klaus Mayer, Eric Richards
                and Steve Rounsley for helpful assistance. In addition, we would
                like to thank the Ohio Supercomputer Center, especially Dixie
                Sky, Troy Ladwig, Michael R. Hanna, Kitty Katsar, Jeremy
                Peterson, Michael Reimer, and Jeffrey Hillard for their assistance and
                database support.

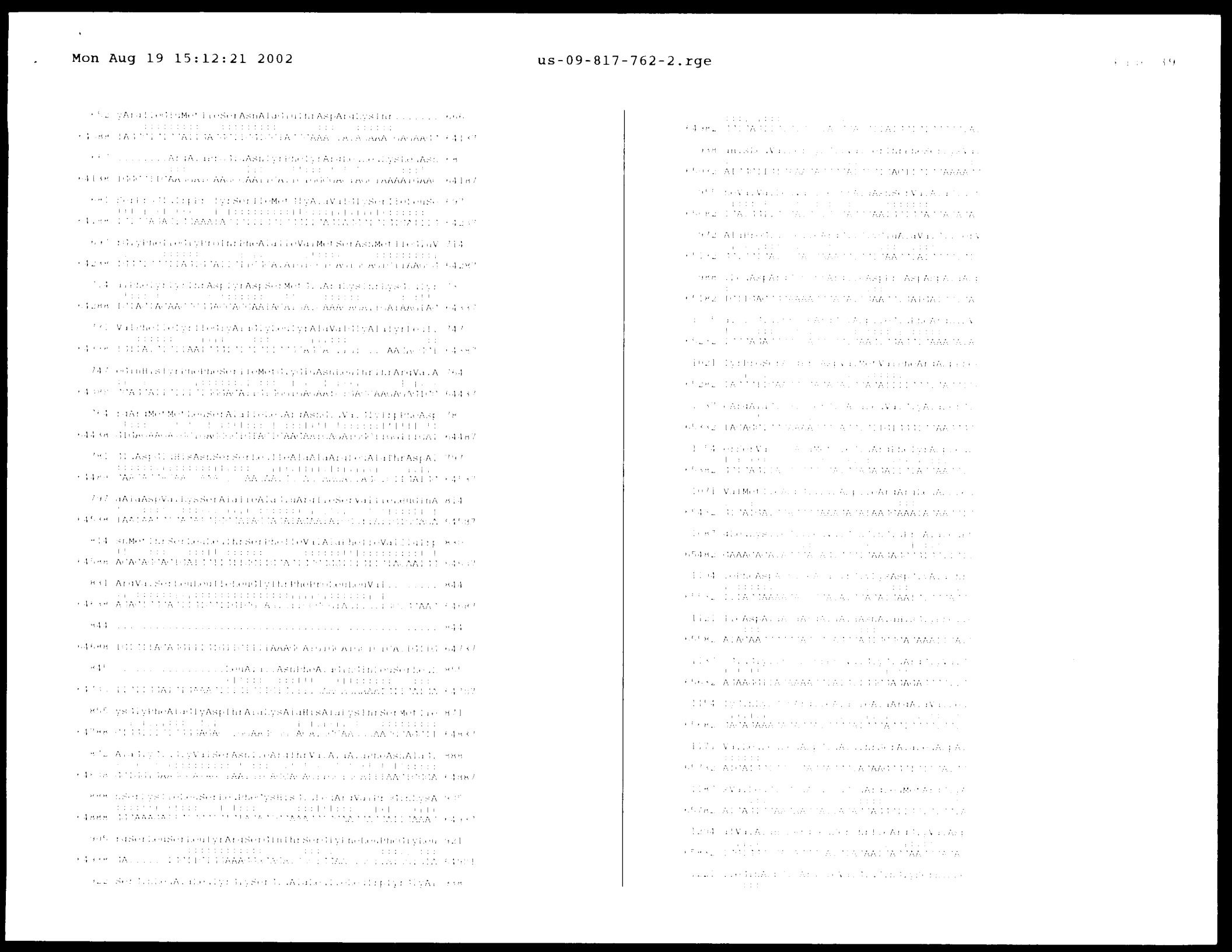
                This work was supported by the National Science Foundation,
                Department of Energy and the US Department of Agriculture.

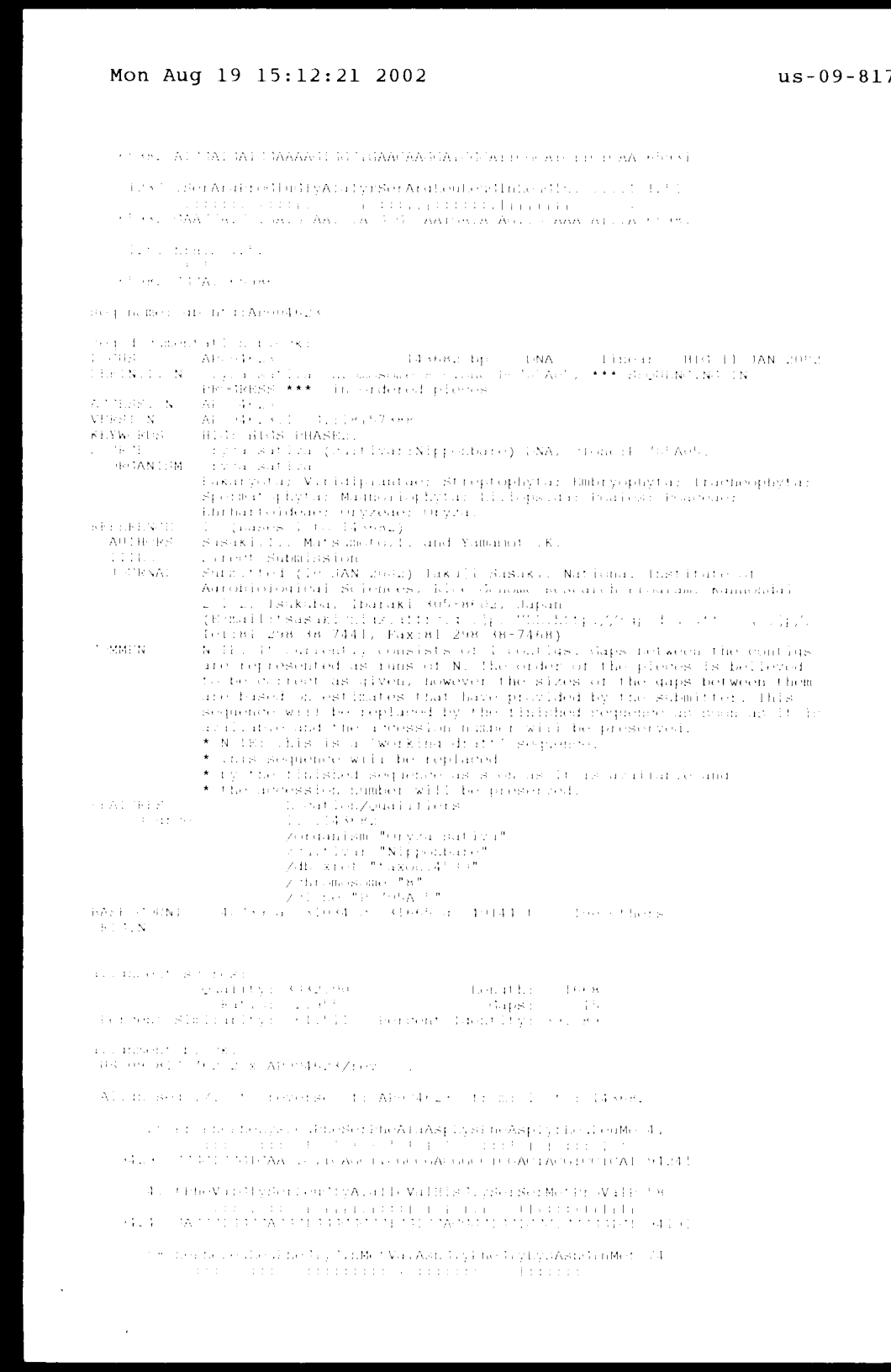
                Address:    Arabidopsis Genome Initiative
FEATURES             Location/Qualifiers
     source          1..119465
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /note="Sequence from clone T13G"

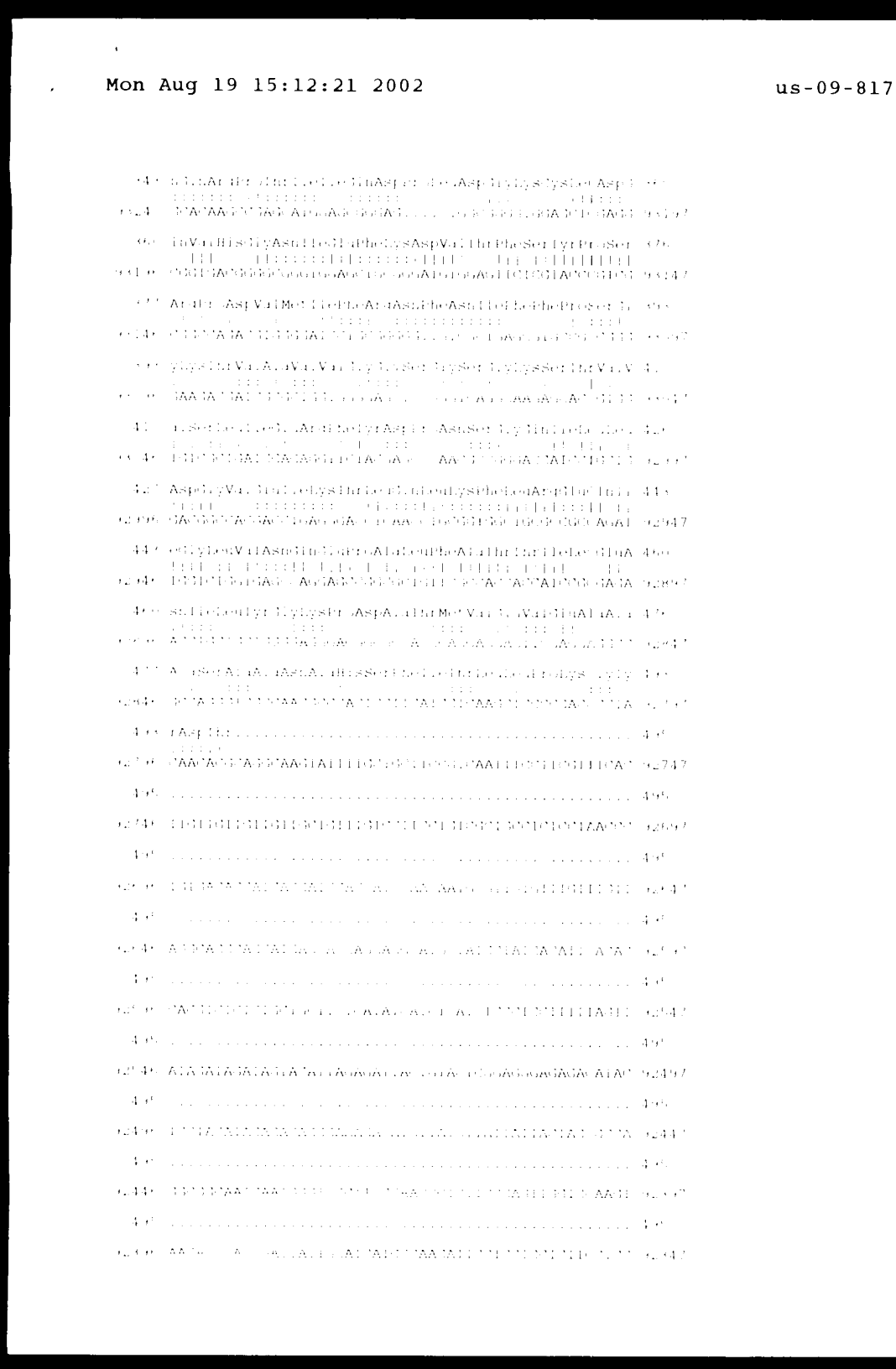
mRNA                 join(complement(join(103840..103967,104066..104146,104467..104566,
                     104671..104767,105453..105581,105681..105846)),
                     complement(join(106830..107007,107066..107146,107467..107566,
                     107671..107767,108453..108581,108681..108846)))
                     /gene="At2g06840"
gene                 complement(join(103840..103967,104066..104146,104467..104566,
                     104671..104767,105453..105581,105681..105846))
                     /gene="At2g06840"
                     /note="F23A5.2; predicted by genefinder"
CDS                  complement(join(103840..103967,104066..104146,104467..104566,
                     104671..104767,105453..105581,105681..105846))
                     /gene="At2g06840"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD15570.1"
                     /db_xref="GI:1663601"
                     /translation="MEIGNIAGRSLEAISLFAGAGGLDLGFVQAGFSIGMAF
                     NKLIPNSAIETSKLVAAYAHDAIFRVAVVGRSVSGKSLPKDNIELLIP
                     SSSWGSLGVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     ASAIIKNSIETAALSLTPVANEIKSLPSFGNNEAVAIDSSVVSKSNSSV
                     YWAGDCVGSTAMEYHVEEINTIEEQIEEIY"
repeat_region        1765..3800
                     /rpt_family="POLY A"
mRNA                 complement(join(103840..103967,104066..104146,104467..104566,
                     104671..104767,105453..105581,105681..105846))
                     /gene="At2g06840"
gene                 complement(103840..105846)
                     /gene="At2g06840"
                     /note="F23A5.2; predicted by genefinder"
CDS                  complement(join(103840..103967,104066..104146,104467..104566,
                     104671..104767,105453..105581,105681..105846))
                     /gene="At2g06840"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD15570.1"
                     /db_xref="GI:1663601"
                     /translation="MEIGNIAGRSLEAISLFAGAGGLDLGFILEF
                     SSSWGSLGHSSNVSIAVDNSISSAHILIHIIQPDHKGLIYIDMKCFKCINLICSYGFE
                     FIKLGKNTFIIEFTVSPPRFILLKSKLNALIIELSAKLLSLGVVMMNKGPDTIELV
                     TNIVELPSKCKFQVFIHDIALAEKKIDKCIISALPGNHVIGKEWEVYKVLINEEDSLP
                     IFDSKIEIVVKIDMWE"
mRNA                 complement(join(13656..13829,14304..14443,14644..14746,14814..14916,15451..15590,15684..15784,16432,16434..16536,
                     16696..16738,17418..17447,17755..17864,17955..18018,18243,18645..18805,18884..18964,19102..19244,19417,19520..19614))
                     /gene="At2g06850"
gene                 complement(13656..19614)
                     /gene="At2g06850"
                     /note="T13G10.3"
CDS                  complement(join(13656..13829,14304..14443,14644..14746,14814..14916,15451..15590,15684..15784,16432,16434..16536,
                     16696..16738,17418..17447,17755..17864,17955..18018,18243,18645..18805,18884..18964,19102..19244,19417,19520..19614))
                     /gene="At2g06850"
                     /codon_start=1
                     /product="putative ornithine cyclase"
                     /protein_id="AAD15571.1"
                     /db_xref="GI:1663602"
                     /translation="MEPIHAVELLEWGMKEIEPELSKGKSPKRVSFELVPRSIGI
                     VKSAAANYPKNLPSPPPEEIINGLIMSMSVSGLMVPLDVFLTYYPLVLYSSSF
                     LPSPNELGISILEYLGKTFPDPWENLLPELDKPESLIAIDLQASSLGALLGDWVSYP
                     GQILAKIVPIMMYYRALMLGCTLEEPKCSVEIAGLLNMPFPLEGSLHAPAADLKEI
                     RVVSGLPLGAAGSKRVMLEAVAPLGSLAVAFPGSLVNGDDSSDPSKK
                     LVKALIDPKKDSELYSKLLWDPKLDCDKPENLNRAIVFFPCPATGLLDMNLCNYLEA
                     SKMLMEIIEDTEKPIFLLLDVPEMHYSHPLVEIPSHLIFLSGKAS
                     EVLNISLGIYAGMPLNSLEMLNLIHHLGLVLNLTALPSGKVAGCNFLVLGPLVVPI
                     NLFDLFRMMSFYLLVVDLYVTDMNPGGNIALLAANLAPLVAKLSGNIALLAANLAPLIVLL
                     IFIAVPMVMGFLLELSLLKAFESFLMQFQLDLPSVFPLFSLGIFCHYPGFLLLFSGAKV
                     YLLGTIAHLVEGFEAAVGLIGFKGVLVEGGELSWESWWEEDQAHIQLLRGKTLELLLS
```

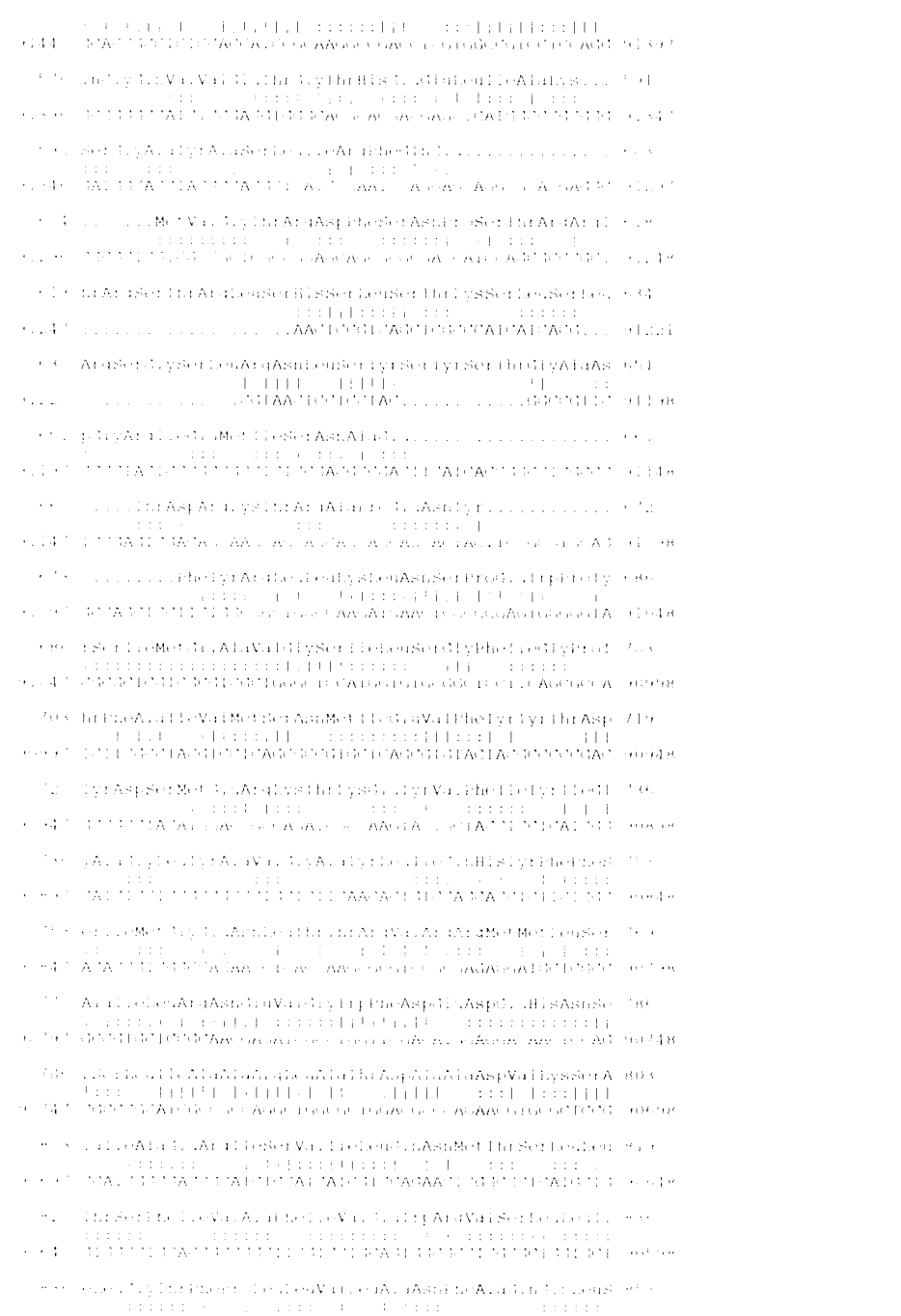


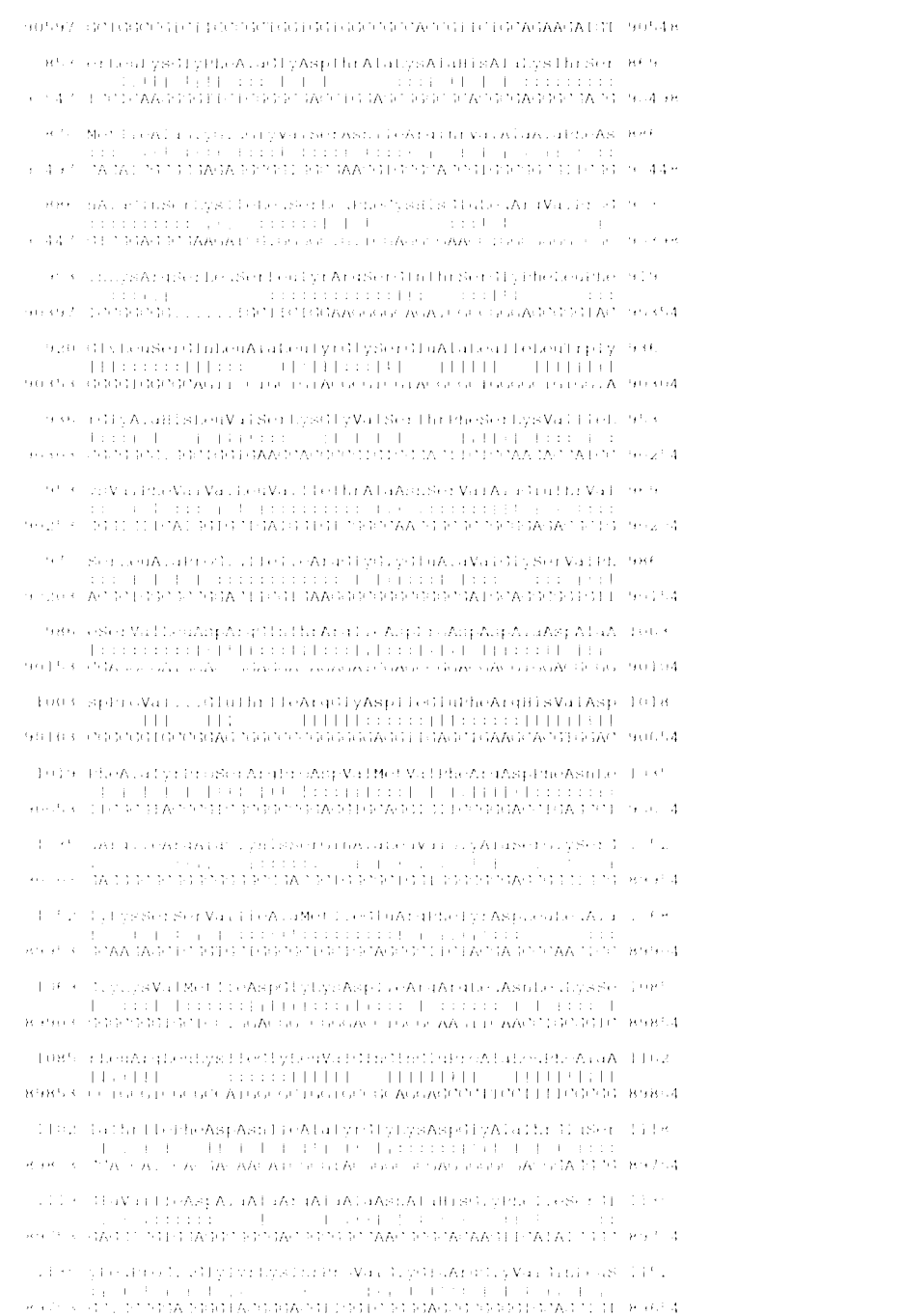


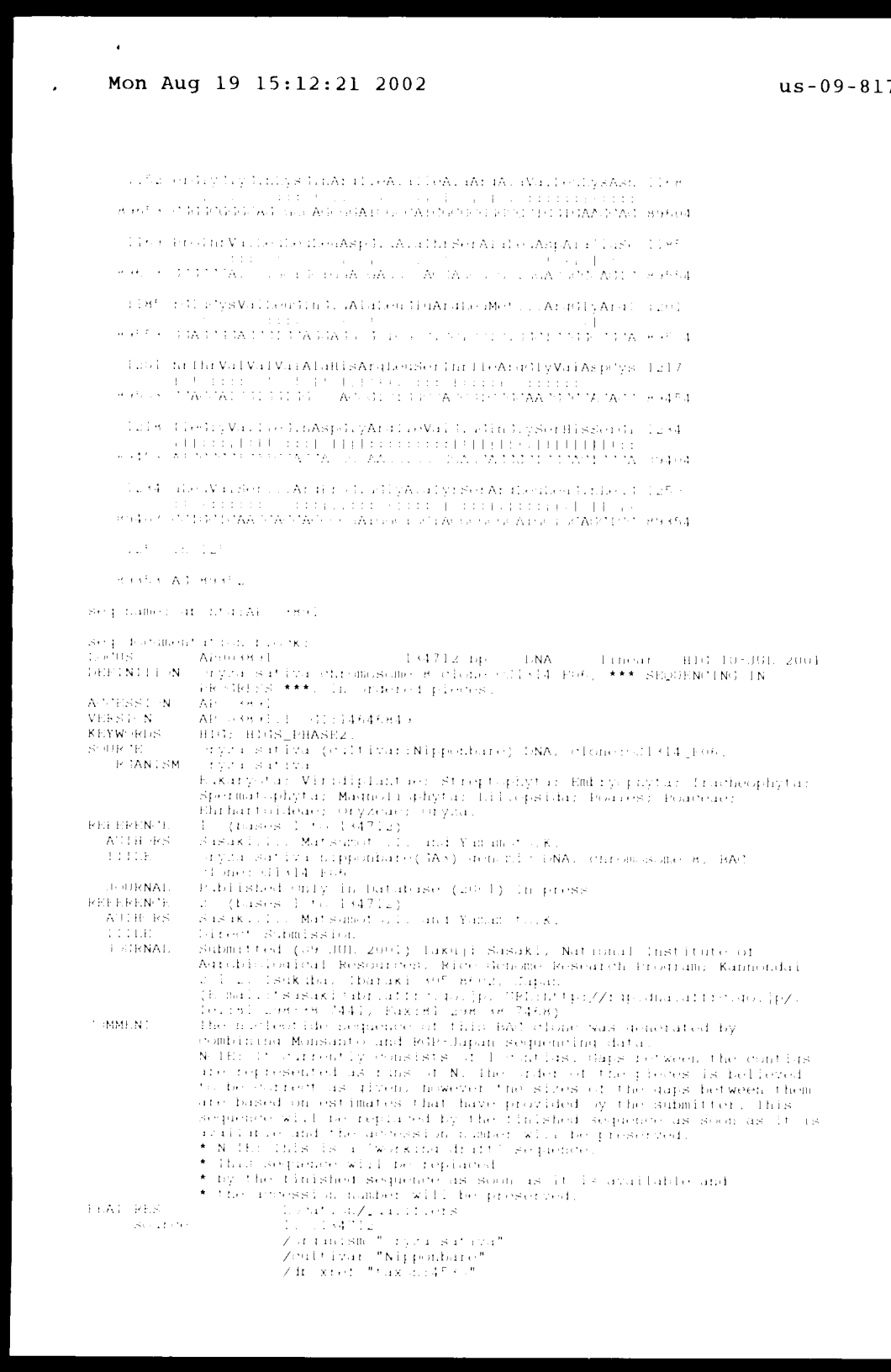








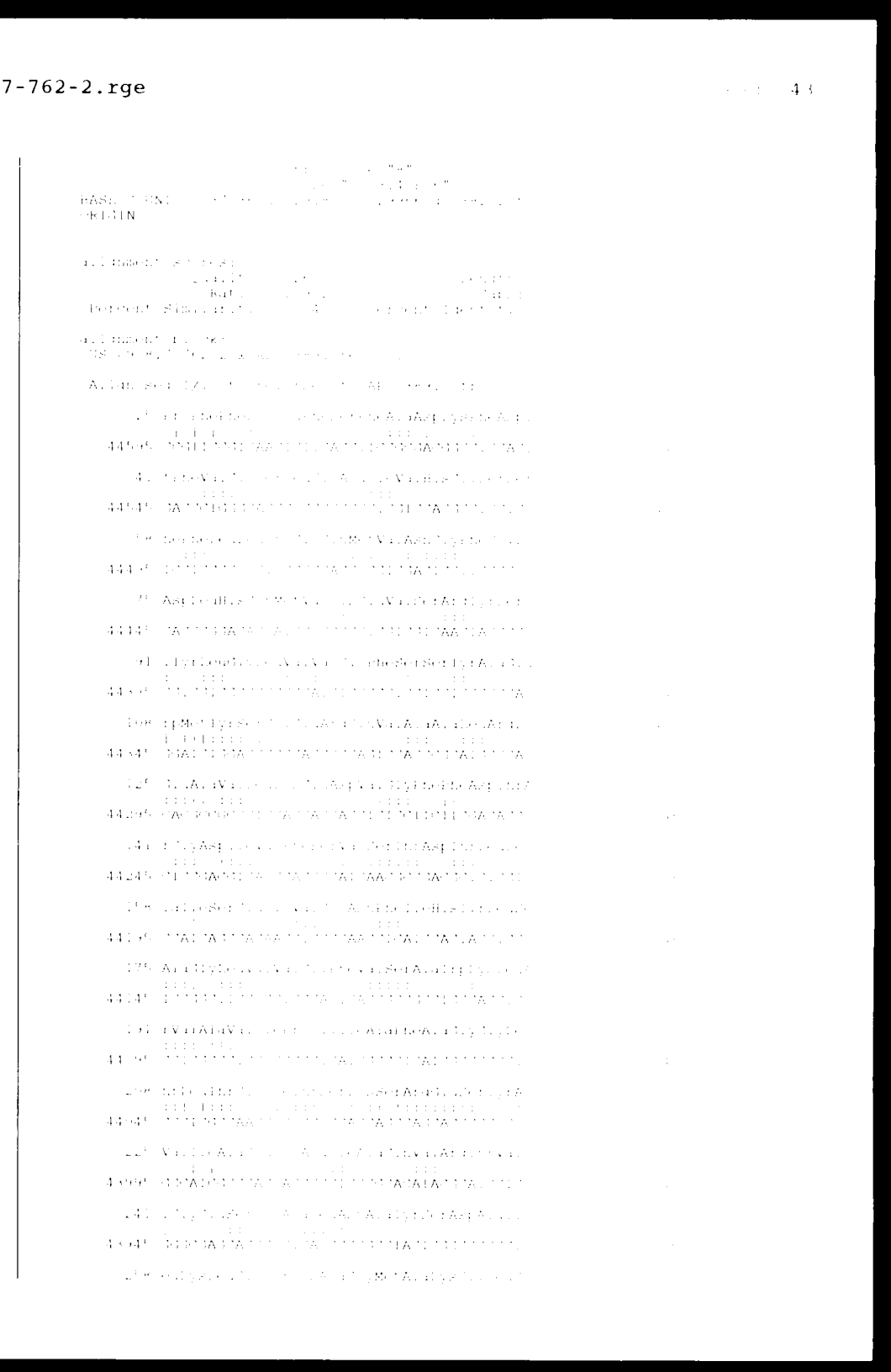


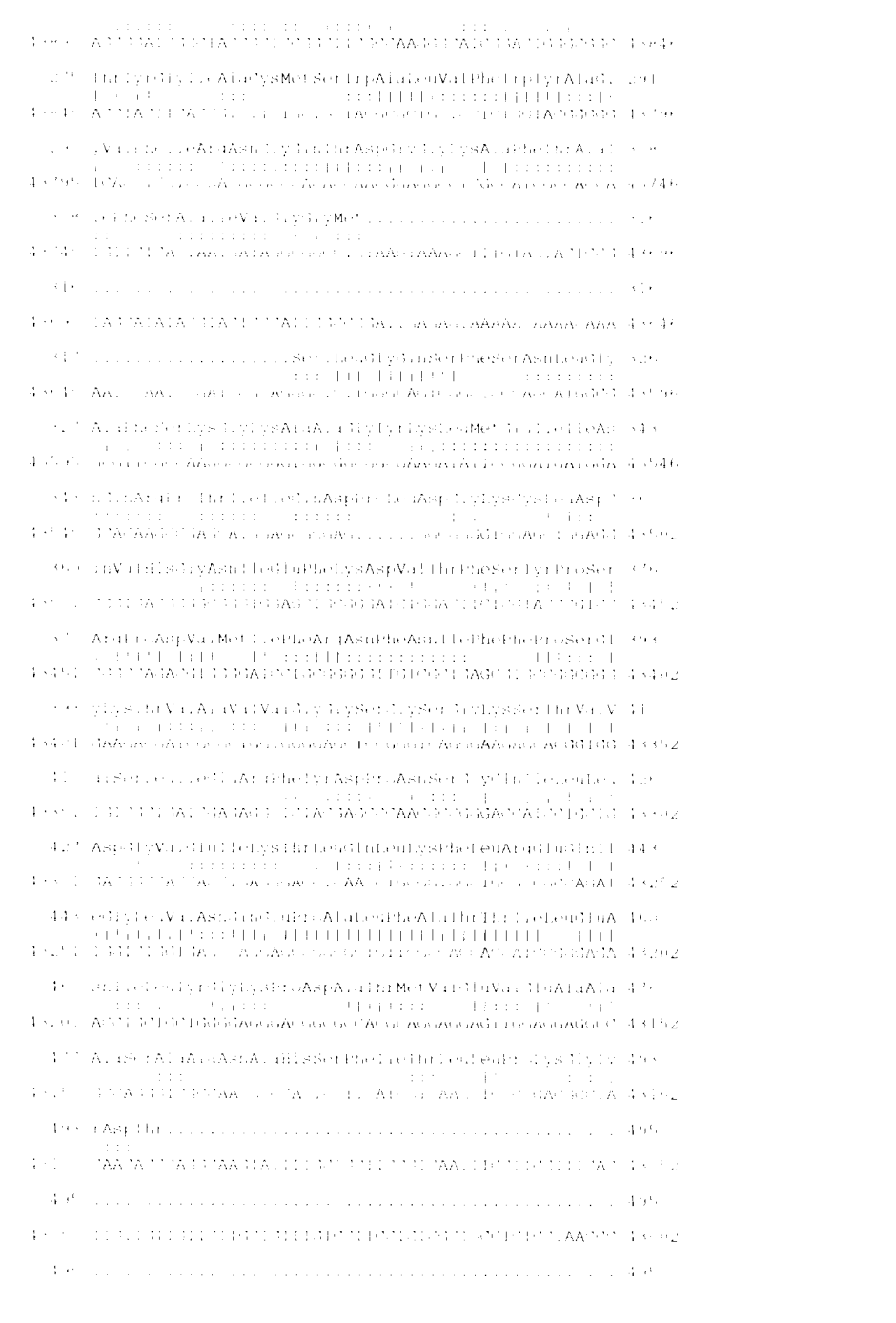


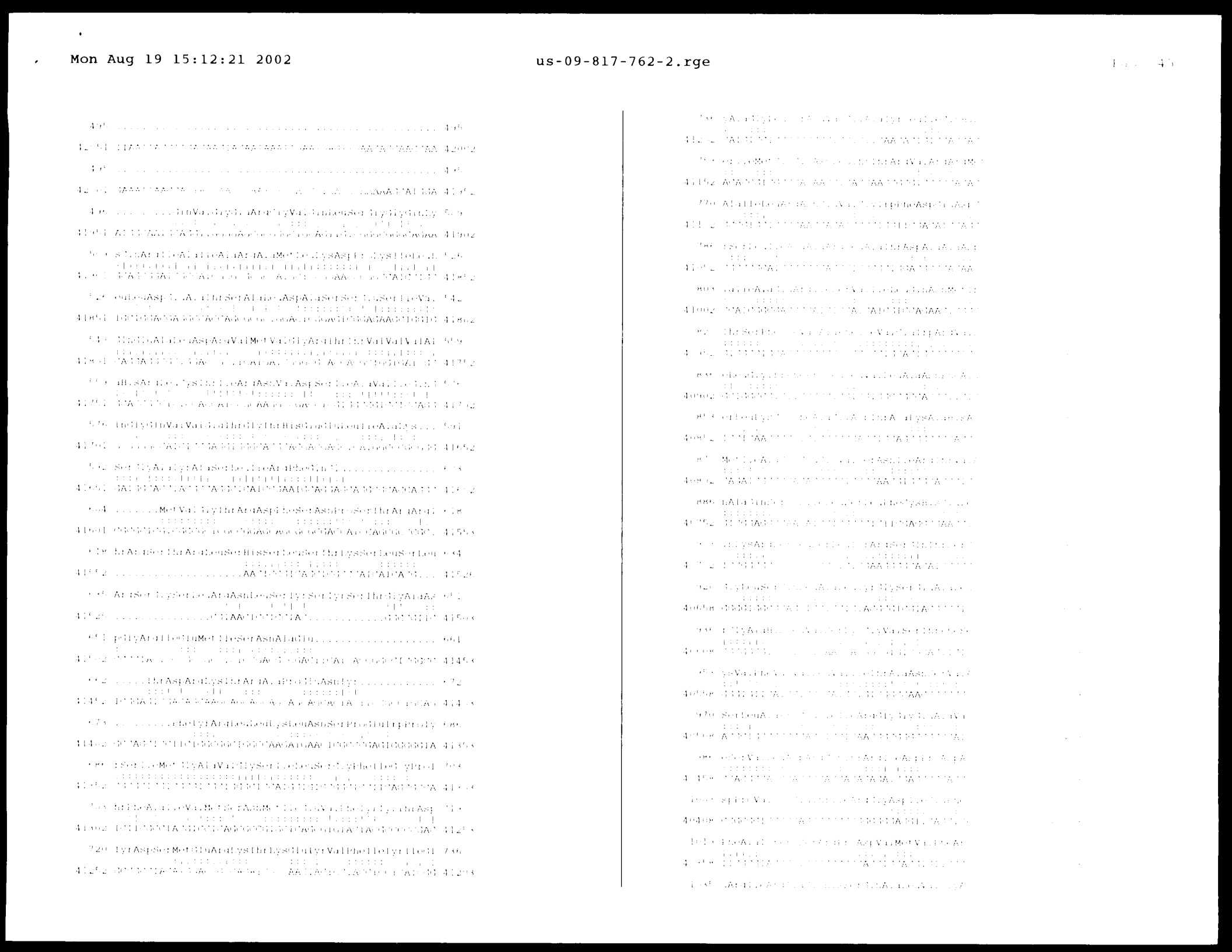
Seq name: ap003144

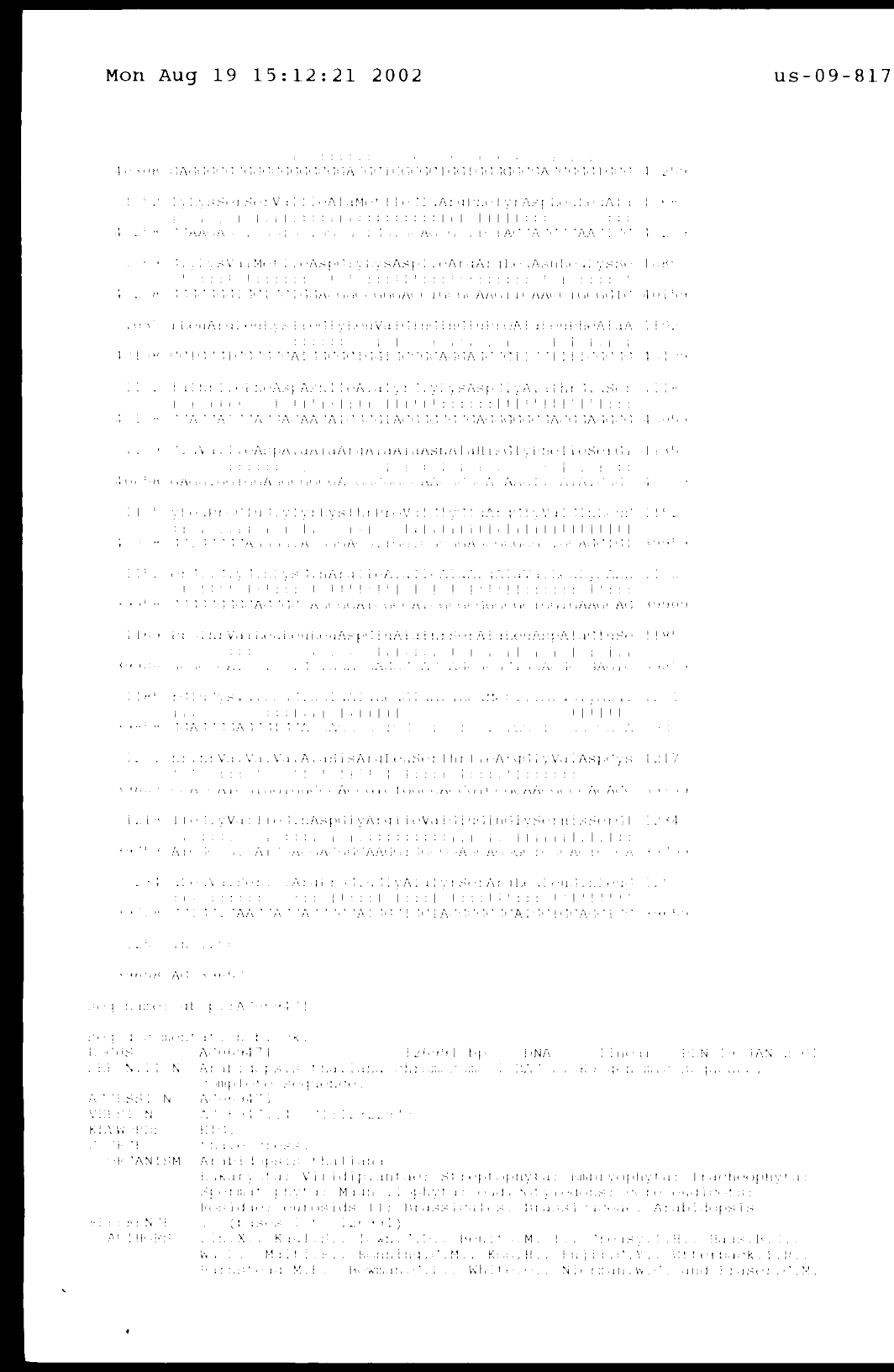
Seq Format: GenBank

locus AP003144 184712 bp DNA linear HTG 10-JUL-2001
DEFINITION Oryza sativa chromosome 8 clone:C1414 F06, *** SEQUENCING IN PROGRESS ***, 56 ordered pieces.
ACCESSION AP003144
VERSION AP003144.1 GI:14464841
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:C1414_F06.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 184712)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(JA) genomic DNA, chromosome 8, BAC clone:C1414 F06
JOURNAL Published only in database (2001) In press
REFERENCE 2 (bases 1 to 184712)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:sasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:+81-298-38-7441, Fax:+81-298-38-7408)
COMMENT The nucleotide sequence of this BAC clone was generated by Combination Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
* NOTE: This is a "working draft" sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES Location/Qualifiers
source 1..184712
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"



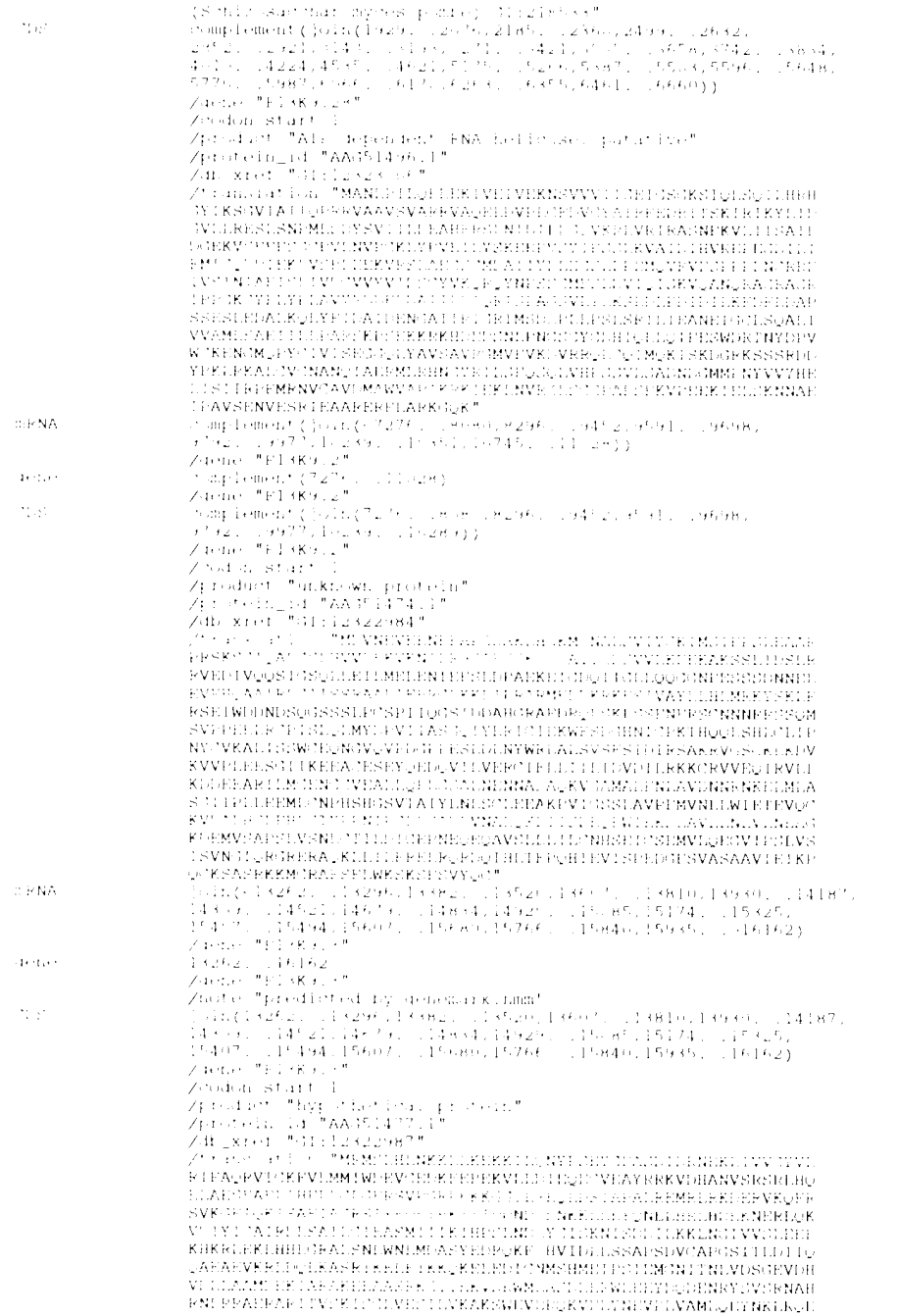


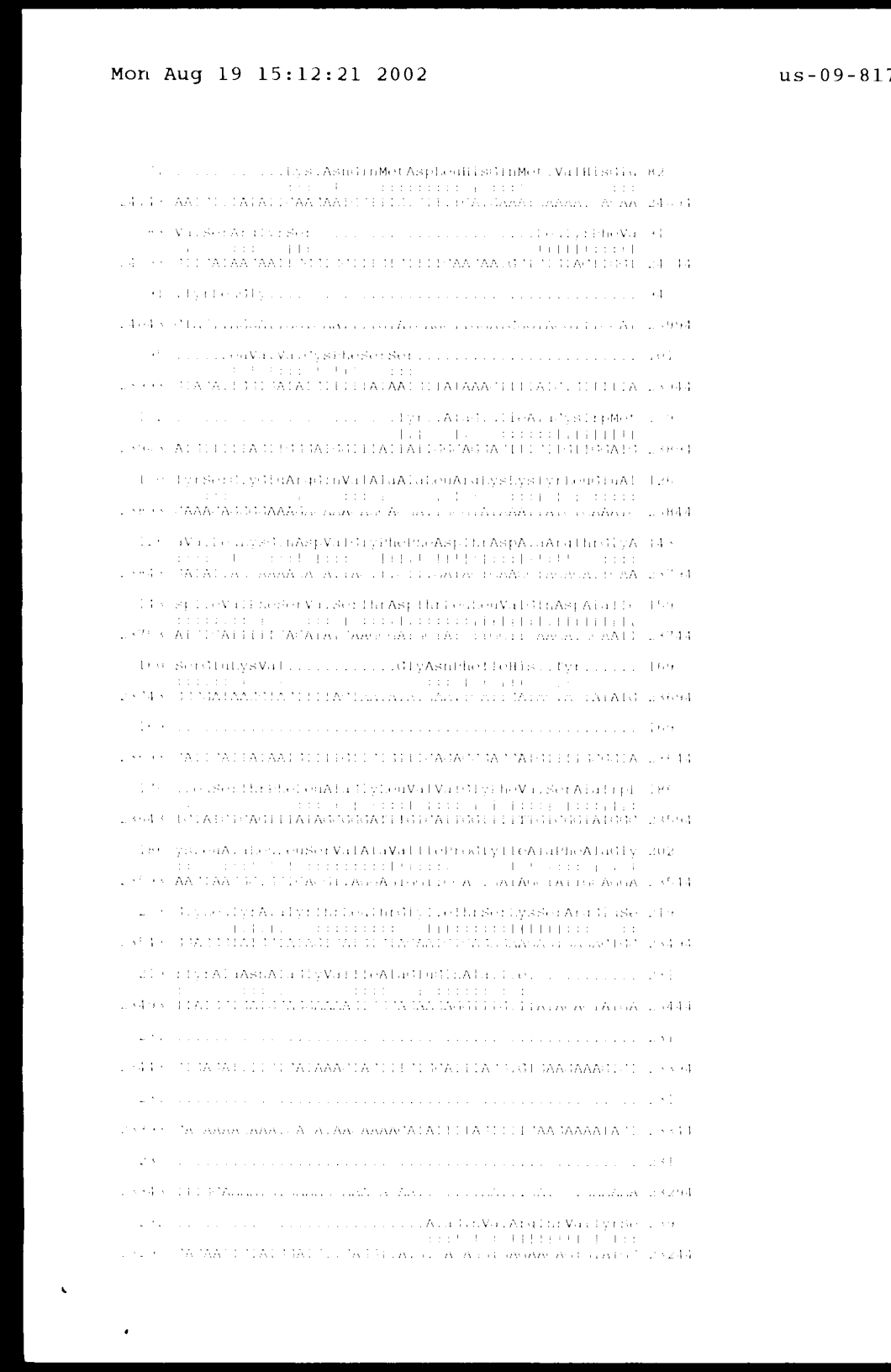


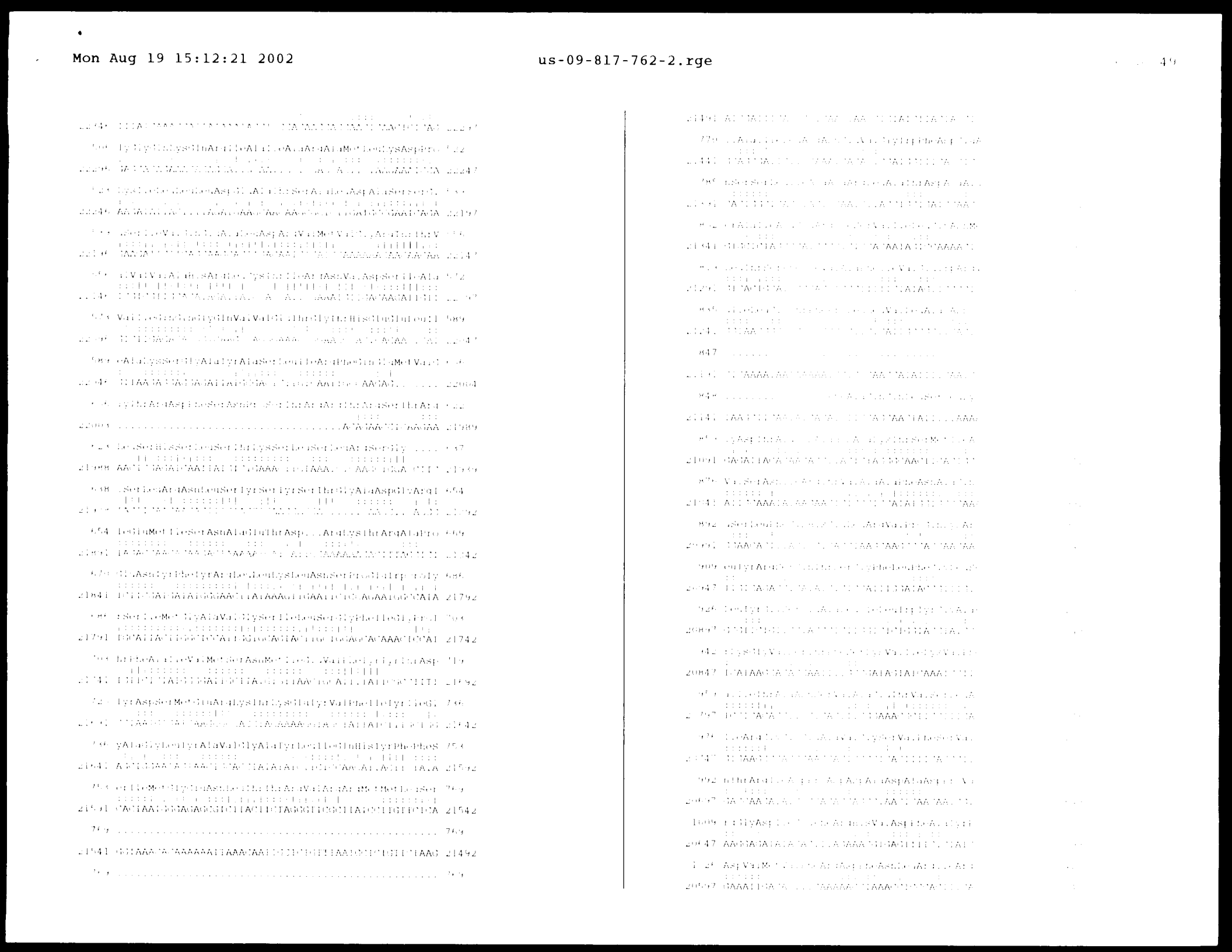


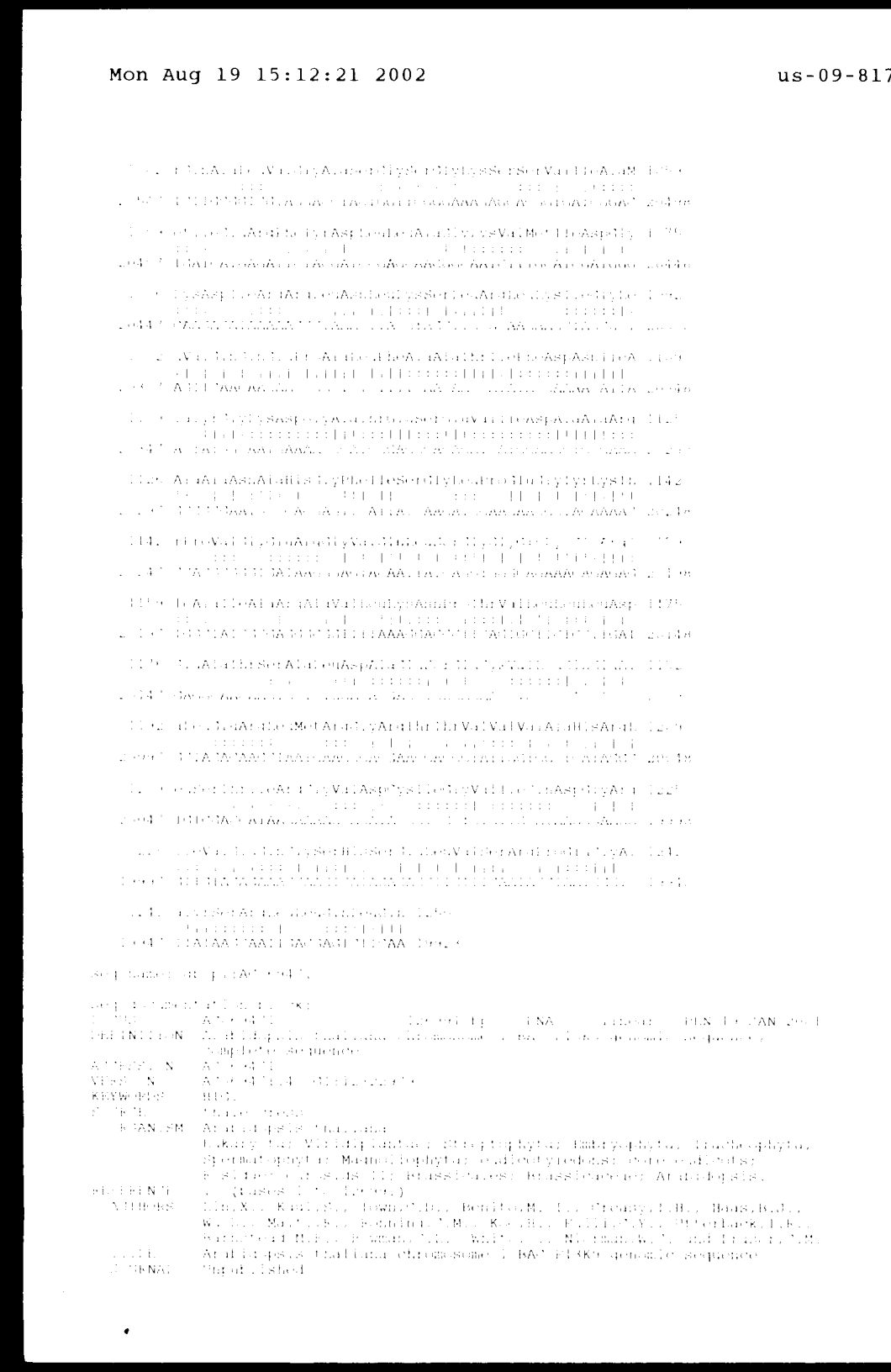
LOCUS Arabidopsis thaliana chromosome 1 BAC F14K9 genomic sequence
DEFINITION mRNA; complete.
ACCESSION AC007591
VERSION AC007591.1 GI:4539477
KEYWORDS HTG; HTG3; high-throughput genomic sequencing.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
REFERENCE 1 (bases 1 to 126991)
AUTHORS Choi,J., Lee,S., Choi,S., Kim,H., Hong,C., Kim,Y., Lee,Y., Seo,H.,
Yang,T., Hwang,I., Choi,I., Kim,M., Cho,S., Kim,S., Kim,K., Park,S.,
Kim,D., Kim,H., Choi,S., Shin,J., Choi,D., Suh,S., Kim,Z., Hong,S.,
Lee,K., Yu,Y., Hur,C., Kim,S., Lee,J., Kim,J., Park,Y., Kim,Z.,
Jin,U., Roe,J., Hong,J., Yi,G., Cheong,G., Arasu,P., Lee,J., Lee,M.,
Park,H., Choi,S., Moon,J., Lee,C., Ko,K., Kim,J., Kim,Y., Kim,T.,
Kim,Y., Kim,S., Kim,W., Kim,H., Moon,Y., Choe,V., Choy,H., Kim,S.,
Nam,J., Oh,B., Kim,K., Cha,S., Chung,M., Nam,M., Lee,Y., Kim,S.,
Noh,J., Cho,K., Kim,H., Kim,J., Kim,B., Kim,S., Baek,S., Choi,B.,
Choi,H., Yang,S., Sohn,S., Hahn,Y., Nam,K., Jekel,J., Yi,S. and
Kim,D.-U.
TITLE Analysis of 110,649 ESTs of Arabidopsis thaliana
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 126991)
AUTHORS Choi,J., Lee,S., Choi,S., Kim,H., Hong,C., Kim,Y., Lee,Y., Seo,H.,
Yang,T., Hwang,I., Choi,I., Kim,M., Cho,S., Kim,S., Kim,K., Park,S.,
Kim,D., Kim,H., Choi,S., Shin,J., Choi,D., Suh,S., Kim,Z., Hong,S.,
Lee,K., Yu,Y., Hur,C., Kim,S., Lee,J., Kim,J., Park,Y., Kim,Z.,
Jin,U., Roe,J., Hong,J., Yi,G., Cheong,G., Arasu,P., Lee,J., Lee,M.,
Park,H., Choi,S., Moon,J., Lee,C., Ko,K., Kim,J., Kim,Y., Kim,T.,
Kim,Y., Kim,S., Kim,W., Kim,H., Moon,Y., Choe,V., Choy,H., Kim,S.,
Nam,J., Oh,B., Kim,K., Cha,S., Chung,M., Nam,M., Lee,Y., Kim,S.,
Noh,J., Cho,K., Kim,H., Kim,J., Kim,B., Kim,S., Baek,S., Choi,B.,
Choi,H., Yang,S., Sohn,S., Hahn,Y., Nam,K., Jekel,J., Yi,S. and
Kim,D.-U.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, mail to: Author
REFERENCE 3 (bases 1 to 126991)
AUTHORS Choi,J., Lee,S., Choi,S., Kim,H., Hong,C., Kim,Y., Lee,Y., Seo,H.,
Yang,T., Hwang,I., Choi,I., Kim,M., Cho,S., Kim,S., Kim,K., Park,S.,
Kim,D., Kim,H., Choi,S., Shin,J., Choi,D., Suh,S., Kim,Z., Hong,S.,
Lee,K., Yu,Y., Hur,C., Kim,S., Lee,J., Kim,J., Park,Y., Kim,Z.,
Jin,U., Roe,J., Hong,J., Yi,G., Cheong,G., Arasu,P., Lee,J., Lee,M.,
Park,H., Choi,S., Moon,J., Lee,C., Ko,K., Kim,J., Kim,Y., Kim,T.,
Kim,Y., Kim,S., Kim,W., Kim,H., Moon,Y., Choe,V., Choy,H., Kim,S.,
Nam,J., Oh,B., Kim,K., Cha,S., Chung,M., Nam,M., Lee,Y., Kim,S.,
Noh,J., Cho,K., Kim,H., Kim,J., Kim,B., Kim,S., Baek,S., Choi,B.,
Choi,H., Yang,S., Sohn,S., Hahn,Y., Nam,K., Jekel,J., Yi,S. and
Kim,D.-U.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, mail to: Author
COMMENT On Jan 19, 2001 this sequence version replaced gi:12207715.
Address all correspondence to the author

BAC F14K9 is from Arabidopsis thaliana chromosome 1.
The orientation of the sequence is from 5' to 3' with T7 at the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan (Chris Burge,
http://CBURGE.MIT.EDU/GENSCAN.HTML), GeneMark.HMM (Mark Borodovsky,
http://exon.gatech.edu/GeneMark/), GlimmerA (Arthur
et GlimmerM, see Mihaela Pertea,
http://www.tigr.org/~mpertea/glimmer/), and
GeneSplicer (Mihaela Pertea and Steven Salzberg, Center
for Bioinformatics, see http://www.cbcb.umd.edu/software/GeneSplicer/). Searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES Location/Qualifiers
source 1..126991
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="m4"
/clone="F14K9"
mRNA complement(<477..>1588)
/gene="F14K9.1"
/note="similar to PRP4 associated factor 1 (Homo sapiens)
SP|Q13435"
CDS complement(<477..>1588)
/gene="F14K9.1"
/codon_start=1
/product="PRP4 associated factor 1 like protein"
/protein_id="AAG51471.1"
/db_xref="GI:12321867"
/translation="MDSVWEWNPEVEMMGTEEGTRHTGTAGLTPGSVTGLPGIPMGAP
EEGAYPFMGTNVHNCHGGLSTGIGGGGTLKIWEINGSSENGHKGKNDKGIAGLKG
NGLNGGKGSEDSGTEPPPPDSAGGKEKKGKPVKTLWVNDQGYPNAYLVKDIGGKPG
EEAGLGHGVVGLGGKLVEVVKKTAKSVGNGPGQGRTAPVGLGMKPVYKAHHAVSG
GEGAPVPVKLGTGGNGKKGVPVVGGGLGVGGLGGMGPGYPGIGPPGPGPGVM
FAAHHPNPYPGGGYPGMP"
tRNA complement(join(<12..>23,<27..>75,<80..>124))
/gene="F14K9.28"
/note="similar to ATP dependent RNA helicase





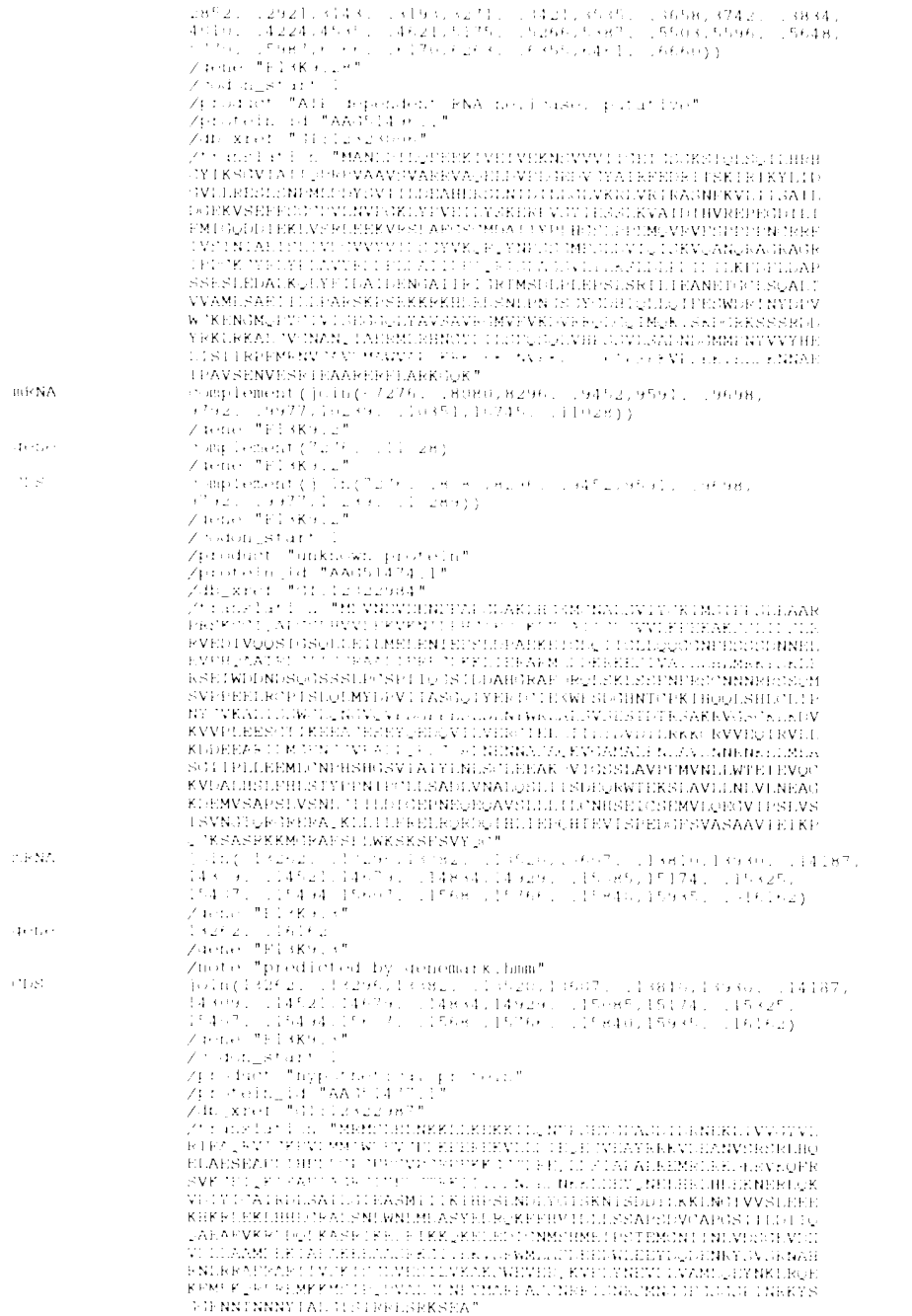


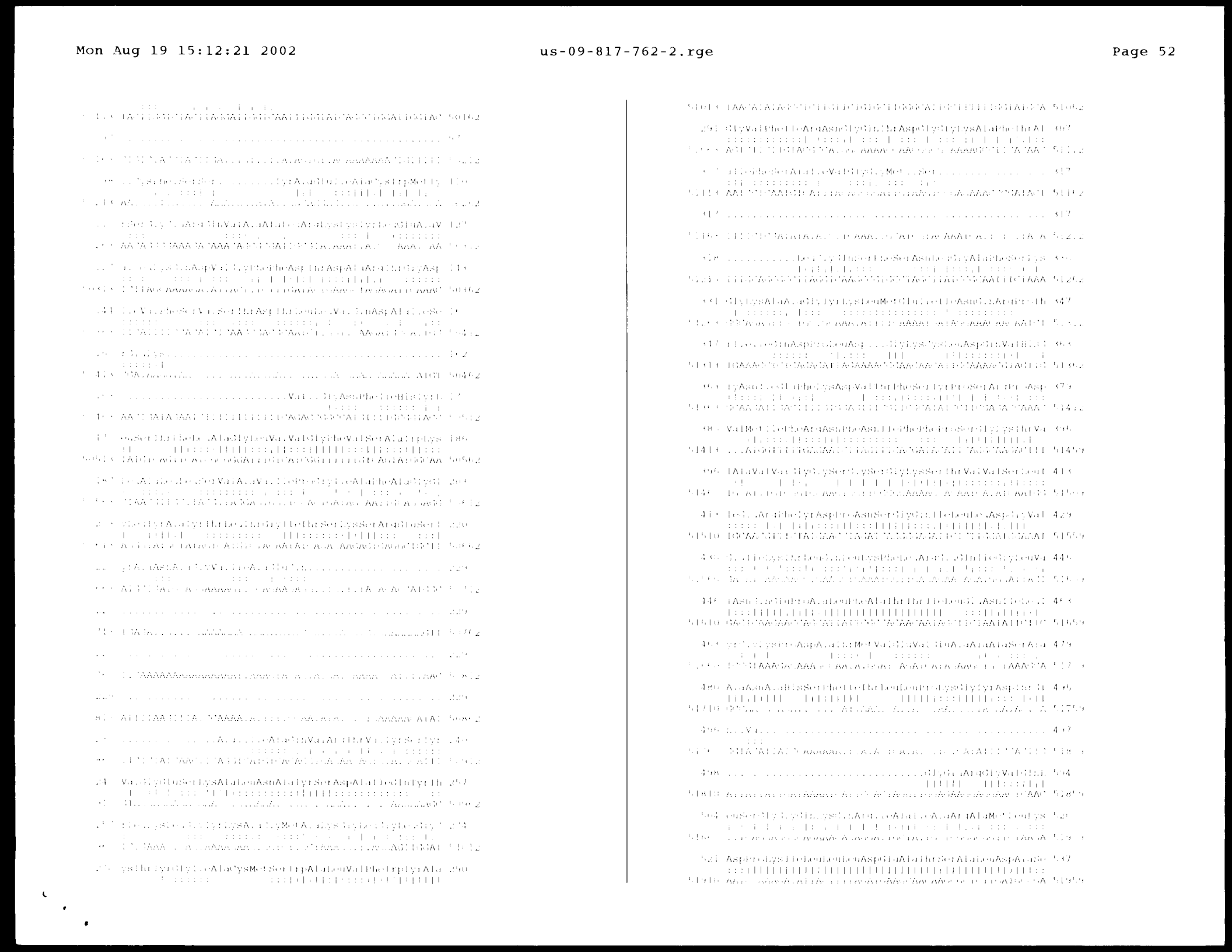


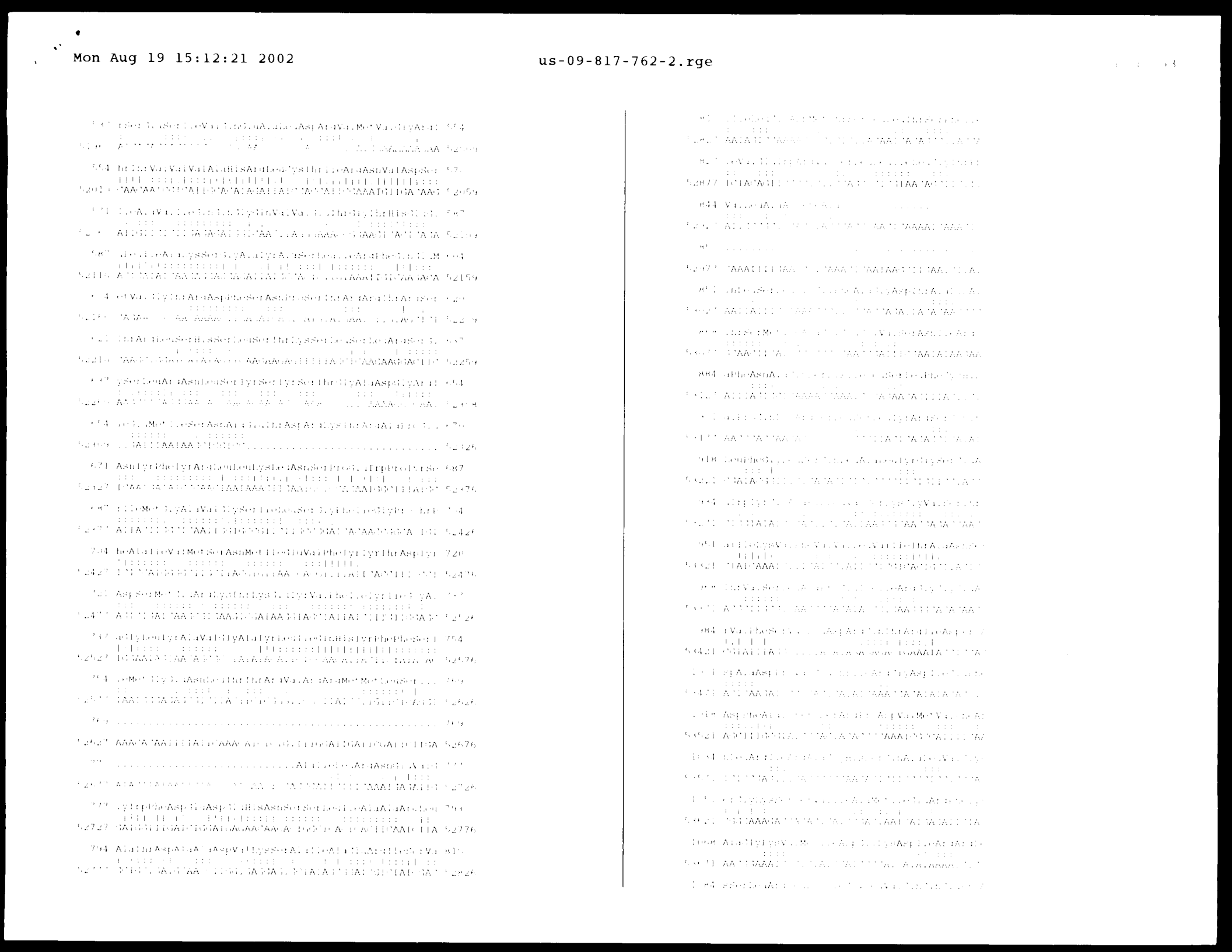
REFERENCE 2 (bases 1 to 126991)
AUTHORS Kaneko,T. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, editor@tigr.org
REFERENCE 3 (bases 1 to 126991)
AUTHORS Kaneko,T. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, editor@tigr.org
COMMENT On Jan 16, 2001 this sequence version replaced gi:12280753.
Address all correspondence to tair@tigr.org

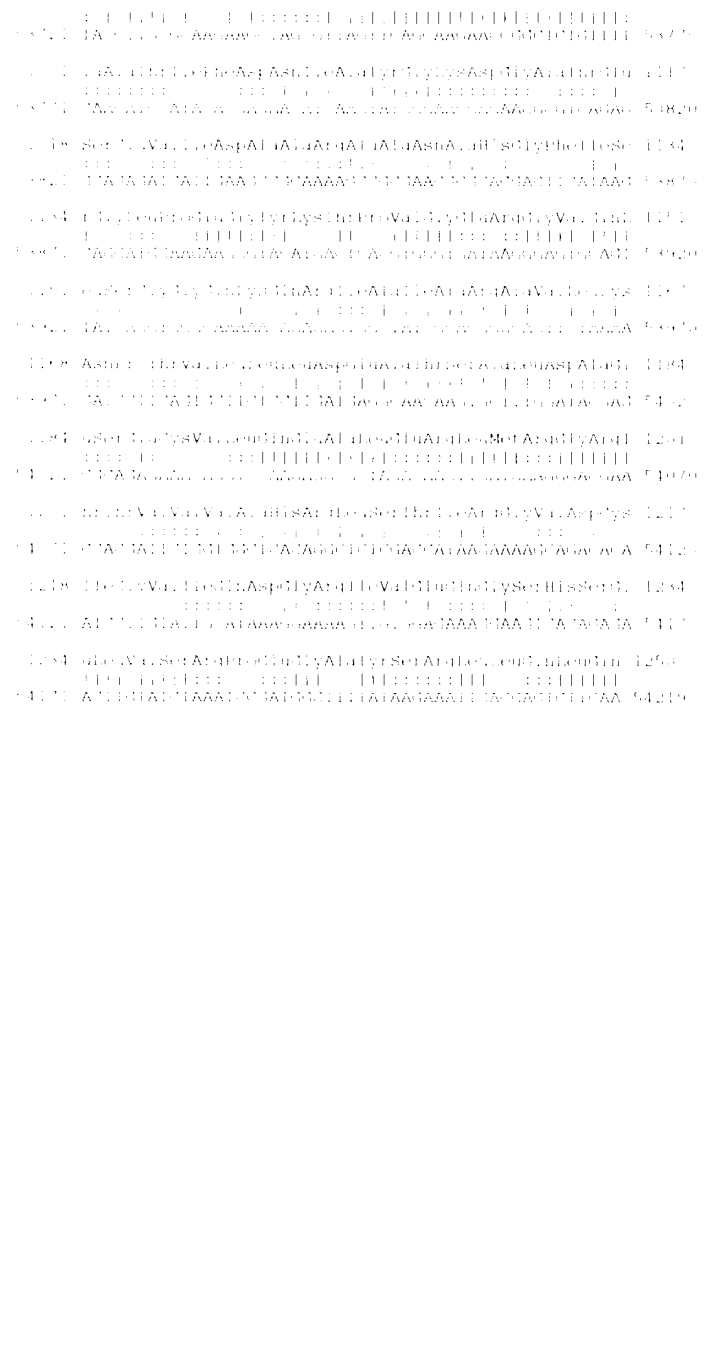
BAC clone is from Arabidopsis thaliana chromosome 2.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of several methods: Gene prediction programs including Genscan (Chris Burge, http://[illegible] [illegible]), GeneMark.hmm (Mark Borodovsky, http://[illegible] [illegible]), GeneFinder (a variant of P. Green's, see Michaels Pertea, http://[illegible] [illegible]), Arithmetic, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact: pertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://[illegible] [illegible]). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown proteins'. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by [illegible] (http://[illegible]). Simple repeats are identified by repeatmasker (Arian Smit, http://[illegible]).

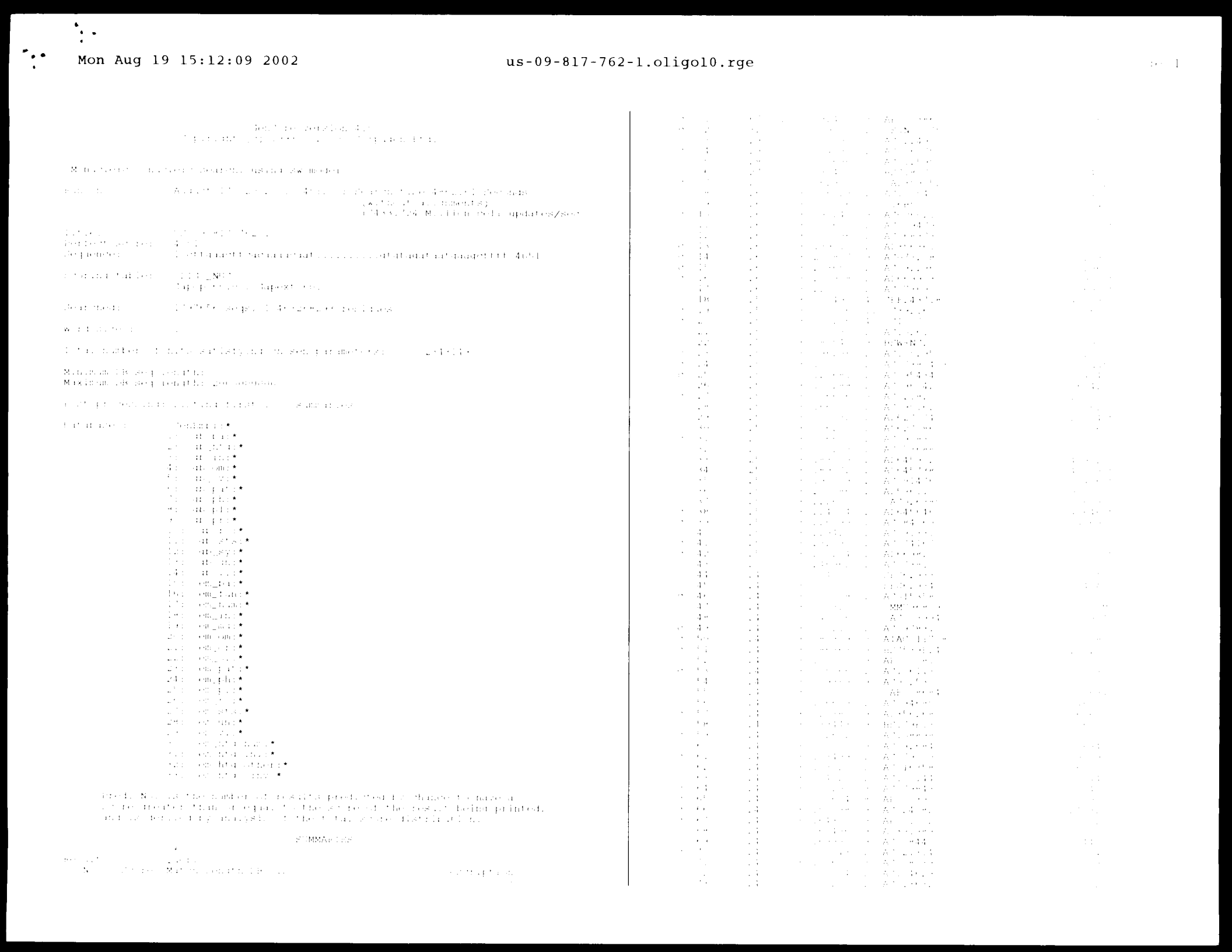
FEATURES Location/Qualifiers
source 1..126991
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="mi-5"
/clone="T12K9"
mRNA complement(<477..>1358)
/gene="F12K9.1"
gene complement(477..1358)
/gene="F12K9.1"
/note="similar to TBP-associated factor 1 (Basic superose, 31:4,p600)"
CDS complement(<477..>1358)
/gene="F12K9.1"
/codon_start=1
/product="TBP-associated factor 1 -like protein, 3' partial"
/protein_id="AAG51123.1"
/db_xref="GI:12322981"
/translation="MLGVWAAAAVTAIIAAAAAAAAAAAAAAAAAAAAMGVG
EEIRVYRDMKFNVDNTHELDIGPILTIRKCIIKIWEINLSEINEHKCTKNGKCTAFIKS
NSLNEIKCGEETCIEEFFECPSQILKEKDGKIIWVNPQGSYFNAYLVKNCLSKKCI
EIKEFFHEIVIQLISKIVEIVKKIALPSSIGSFIGQRIALVLQMKPVGKAHHAICG
SELIARVFIKLIPILENGKQVVRPVGHGGVGVDLGGLDMKGIPVYIPIDPEFPGGVM
FAAHHPNPYPGGGYPGMP"
mRNA complement(join(<1726..2376,2104..2360,4144..4254,
4310..4424,4594..4621,5175..5276,5987..6055,6594..7148,
7211..7607,7664..>7741,8204..8485..>8551))
/note="F12K9.28"
gene complement(1726..6660)
/note="F12K9.28"
/note="similar to AT1 (Arabidopsis) DNA helicase (Schizosaccharomyces pombe) GI:2105733"
CDS complement(join(1926..2076,2184..2360,2499..2944,

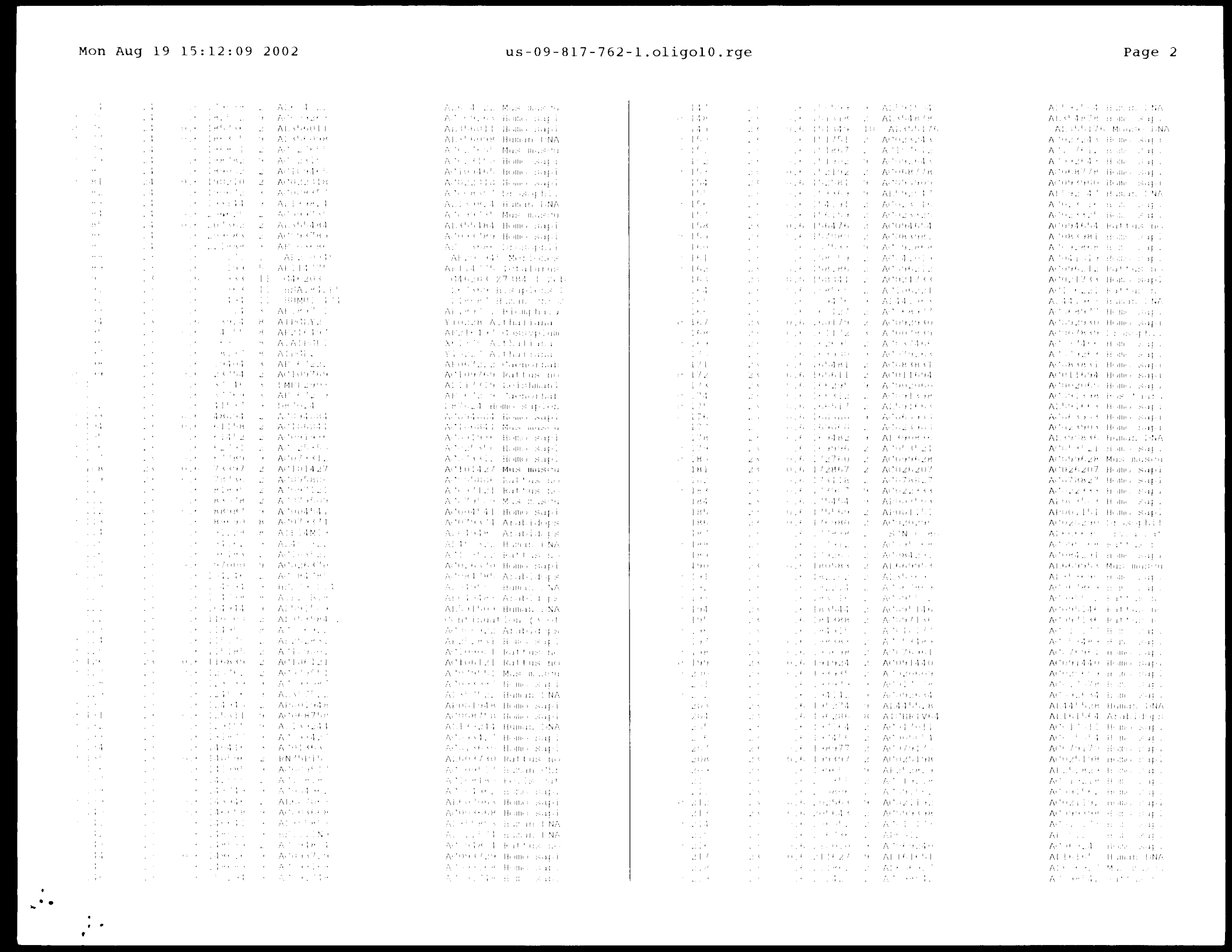


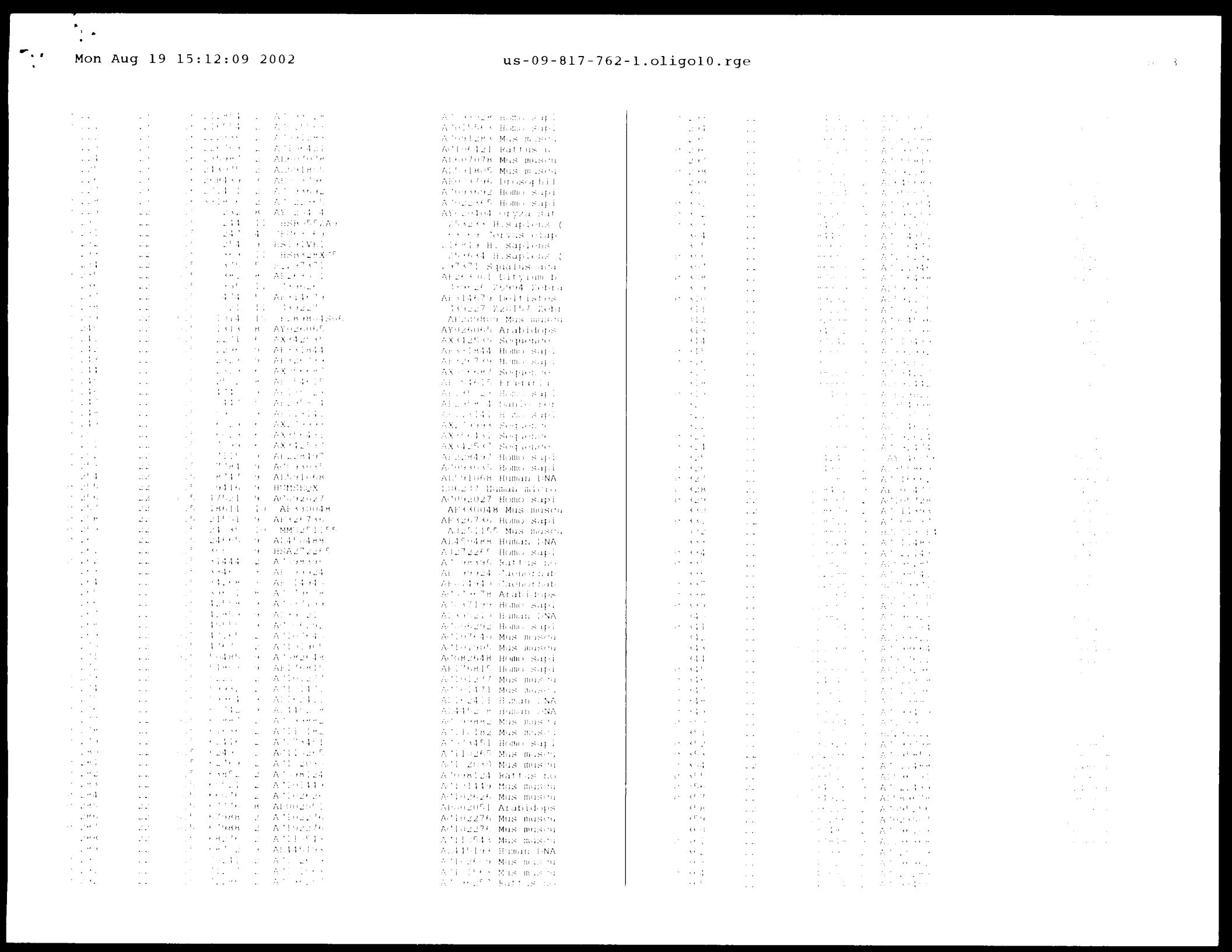


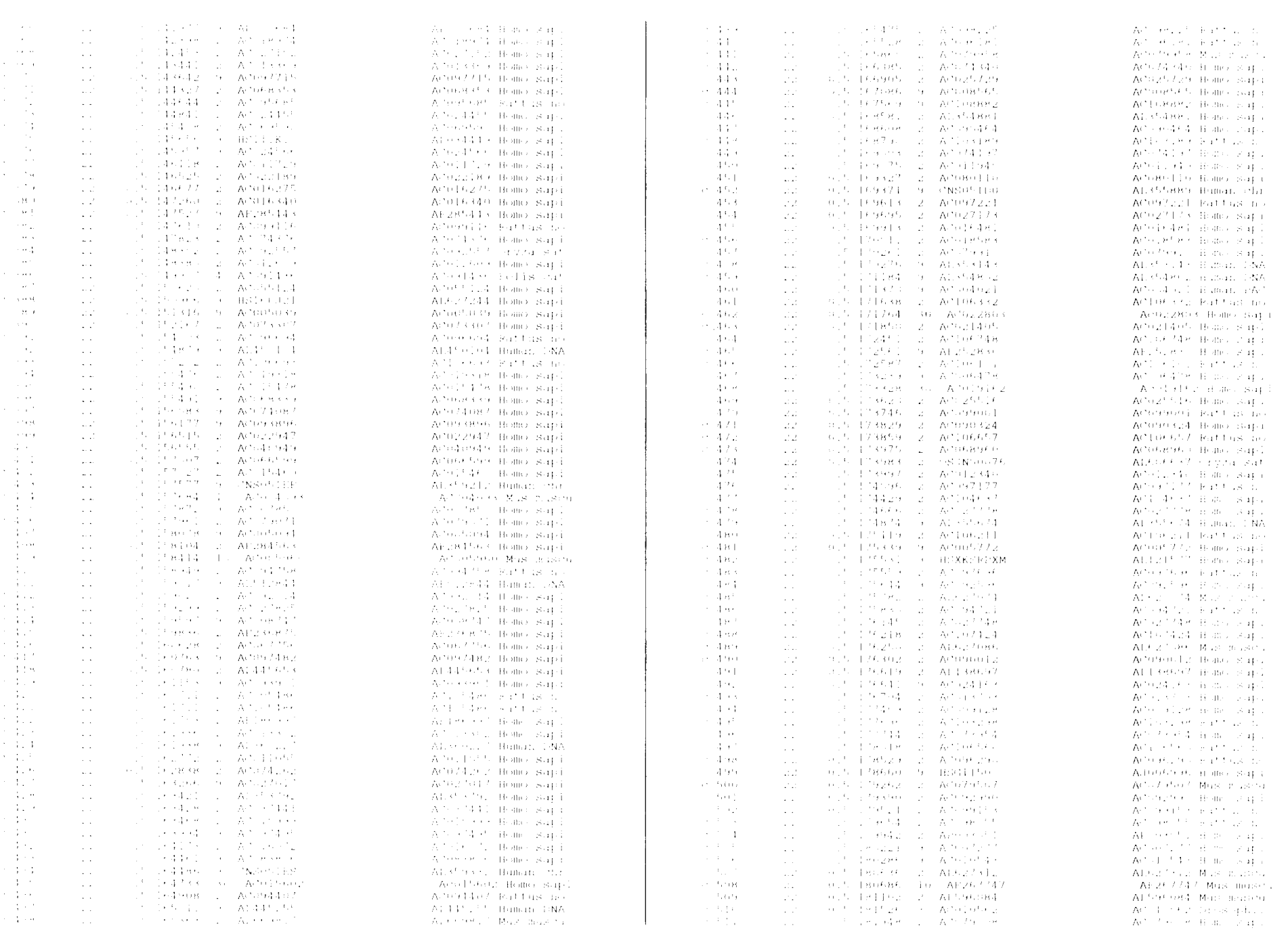


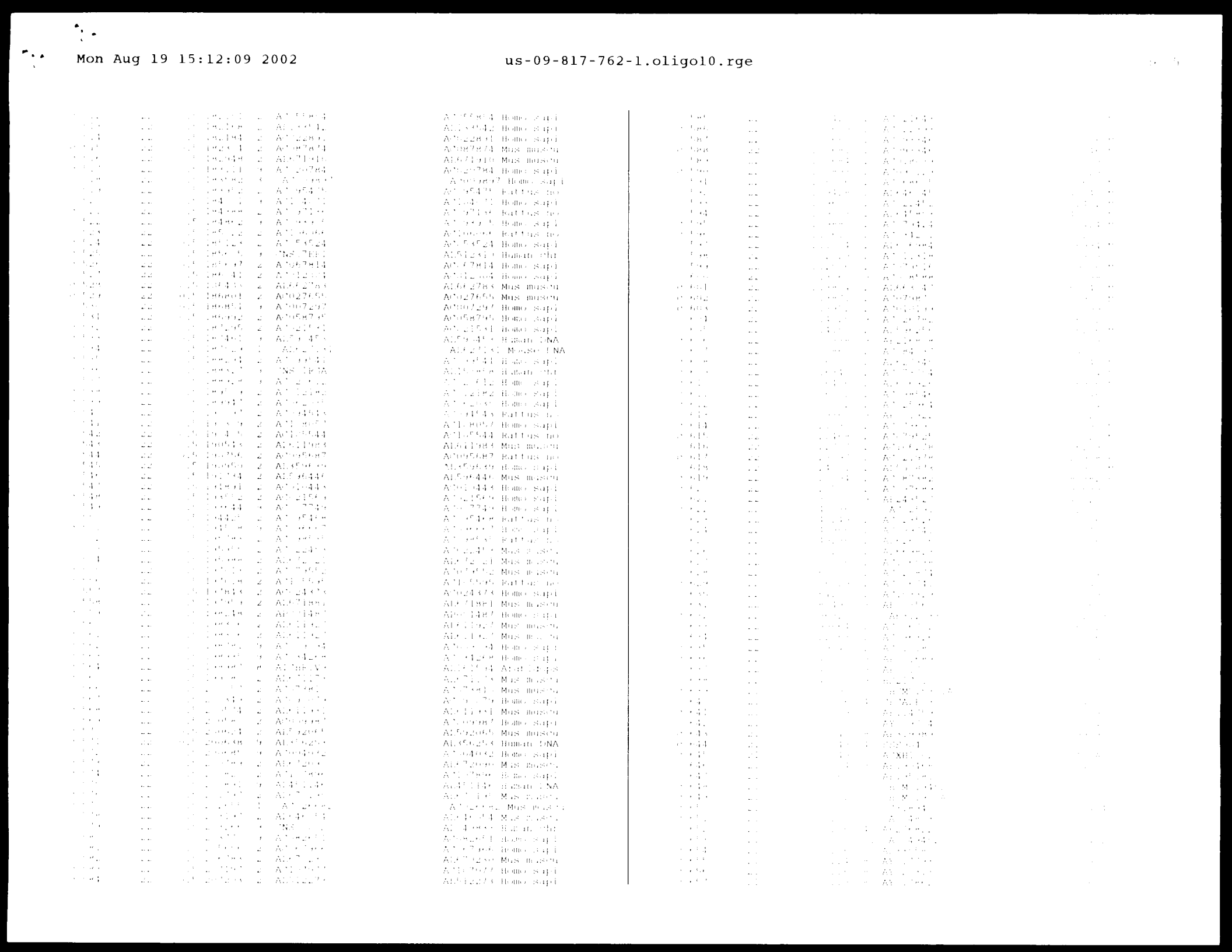


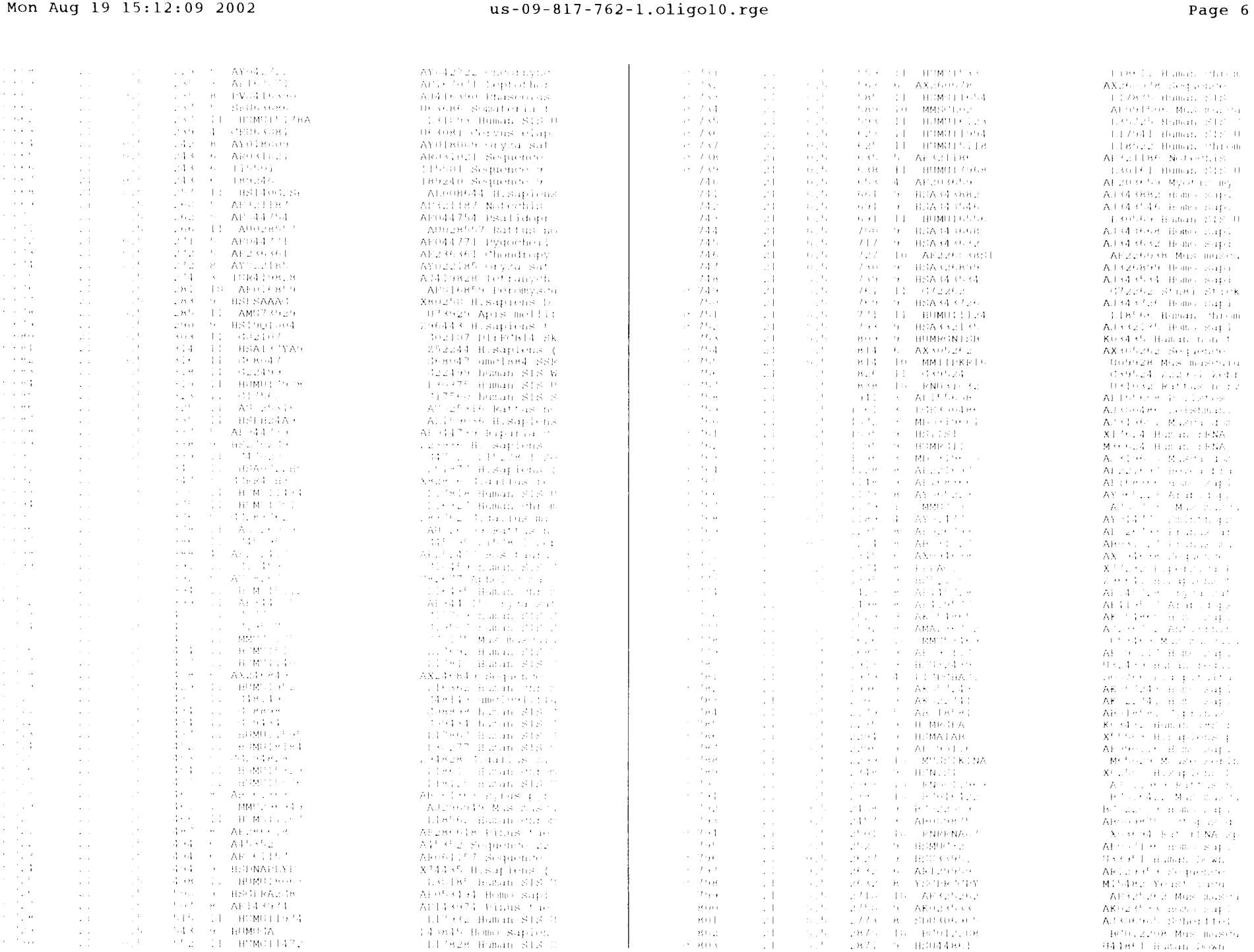


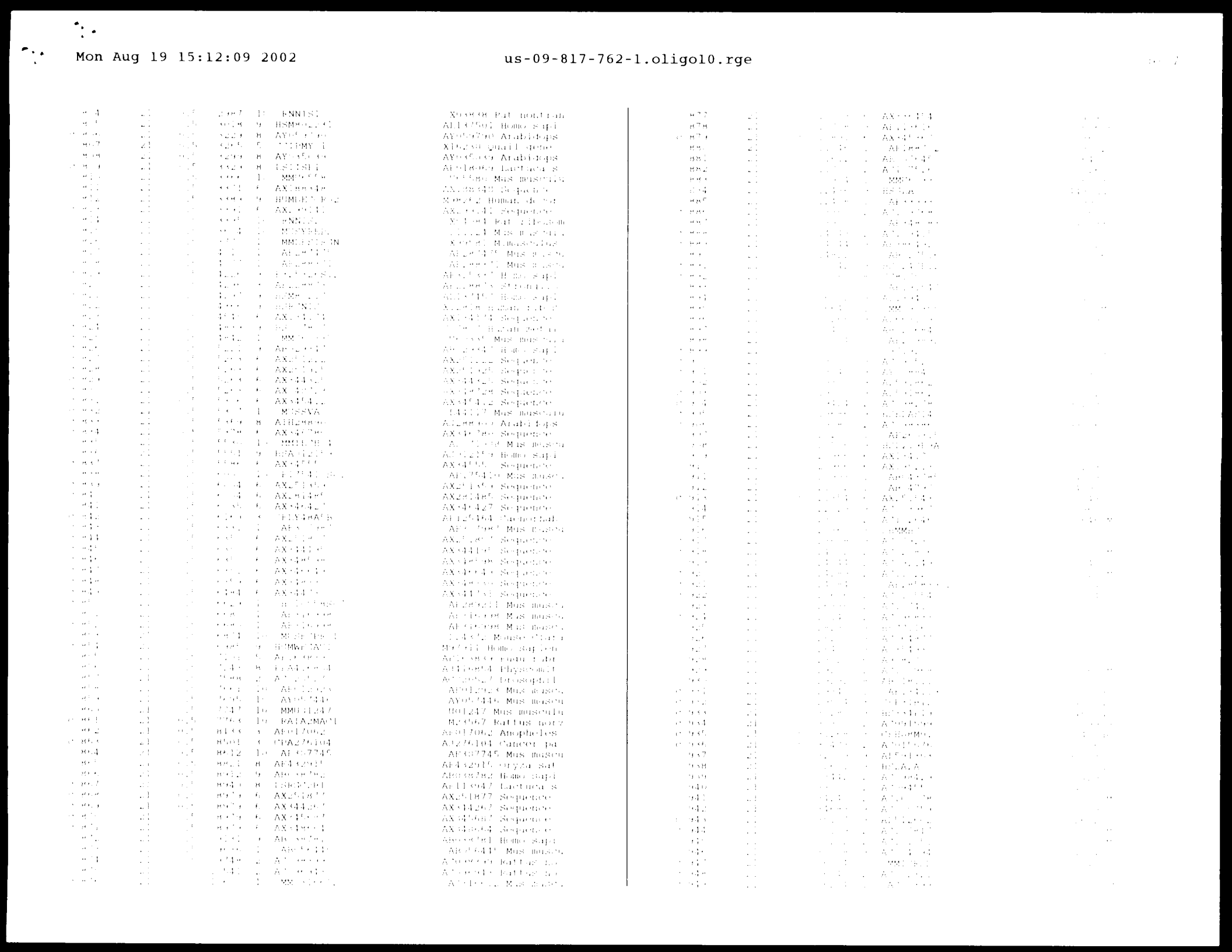


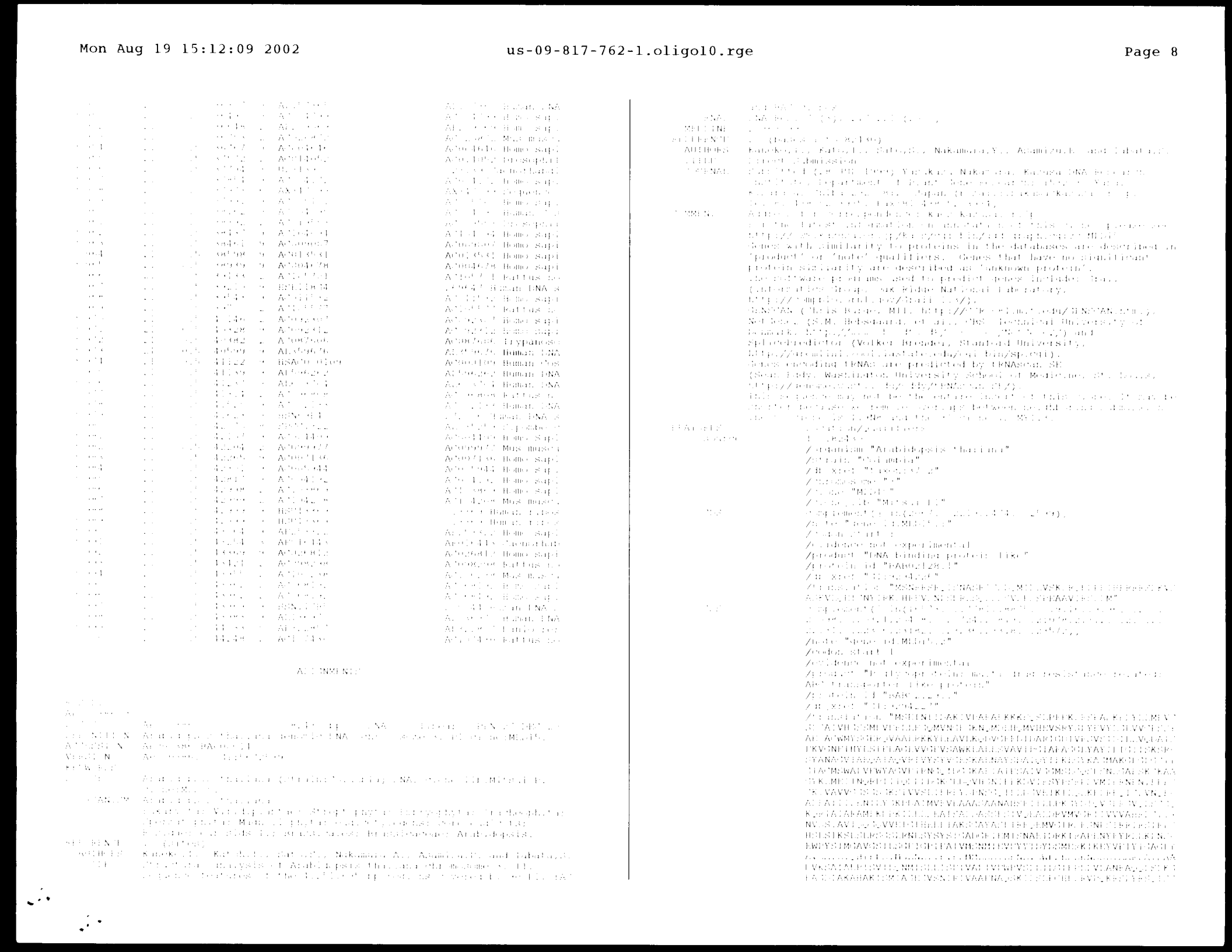


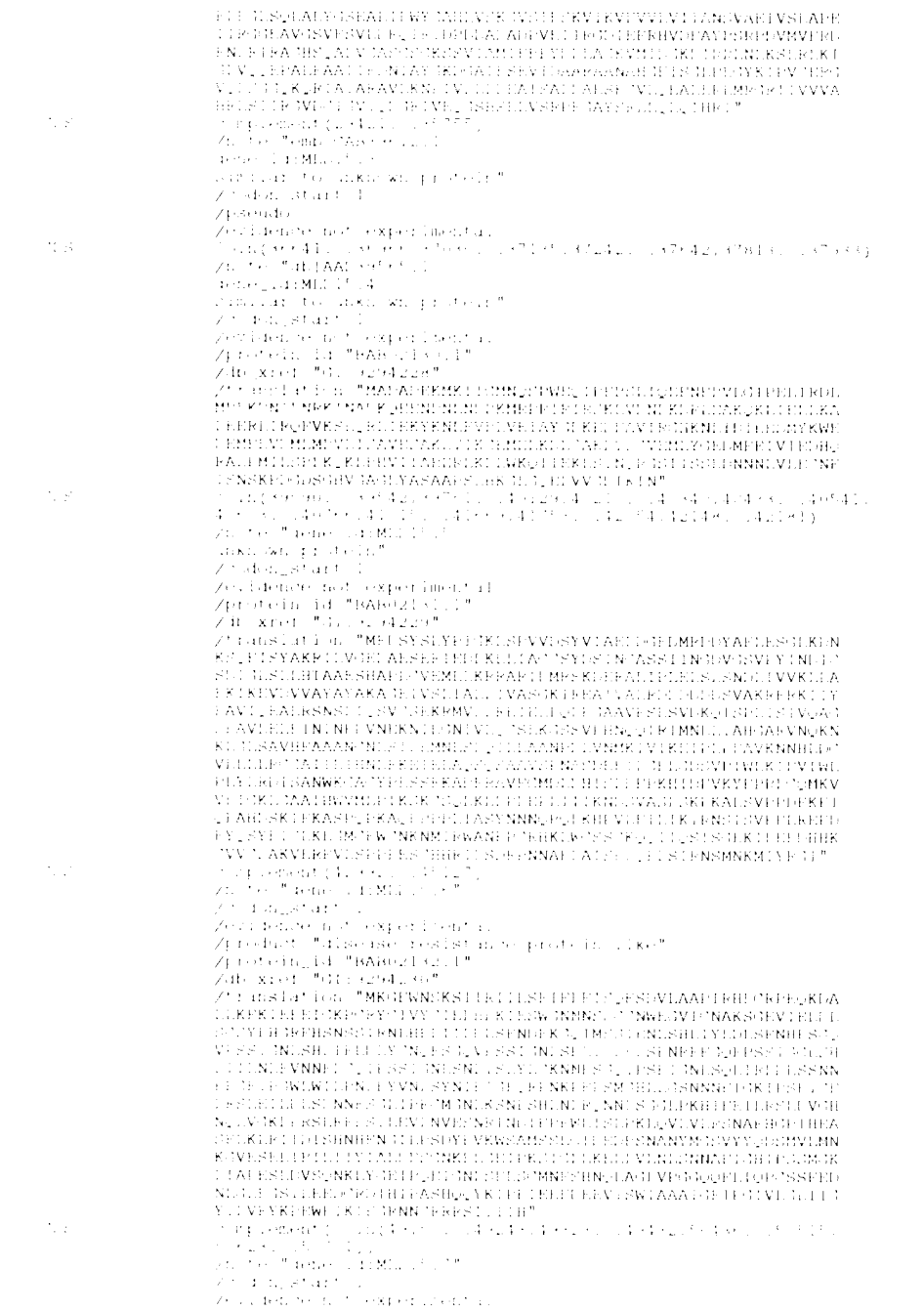


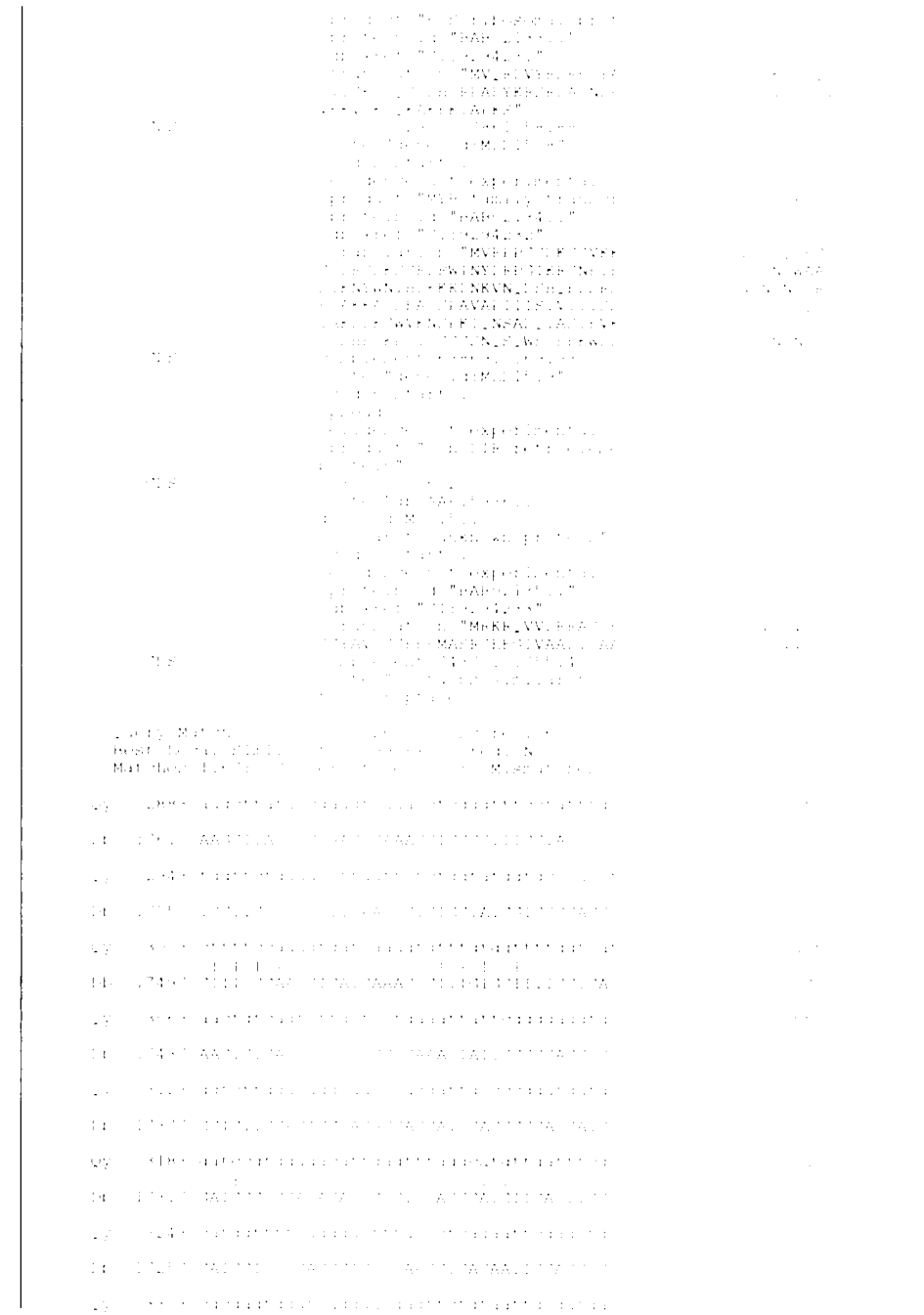


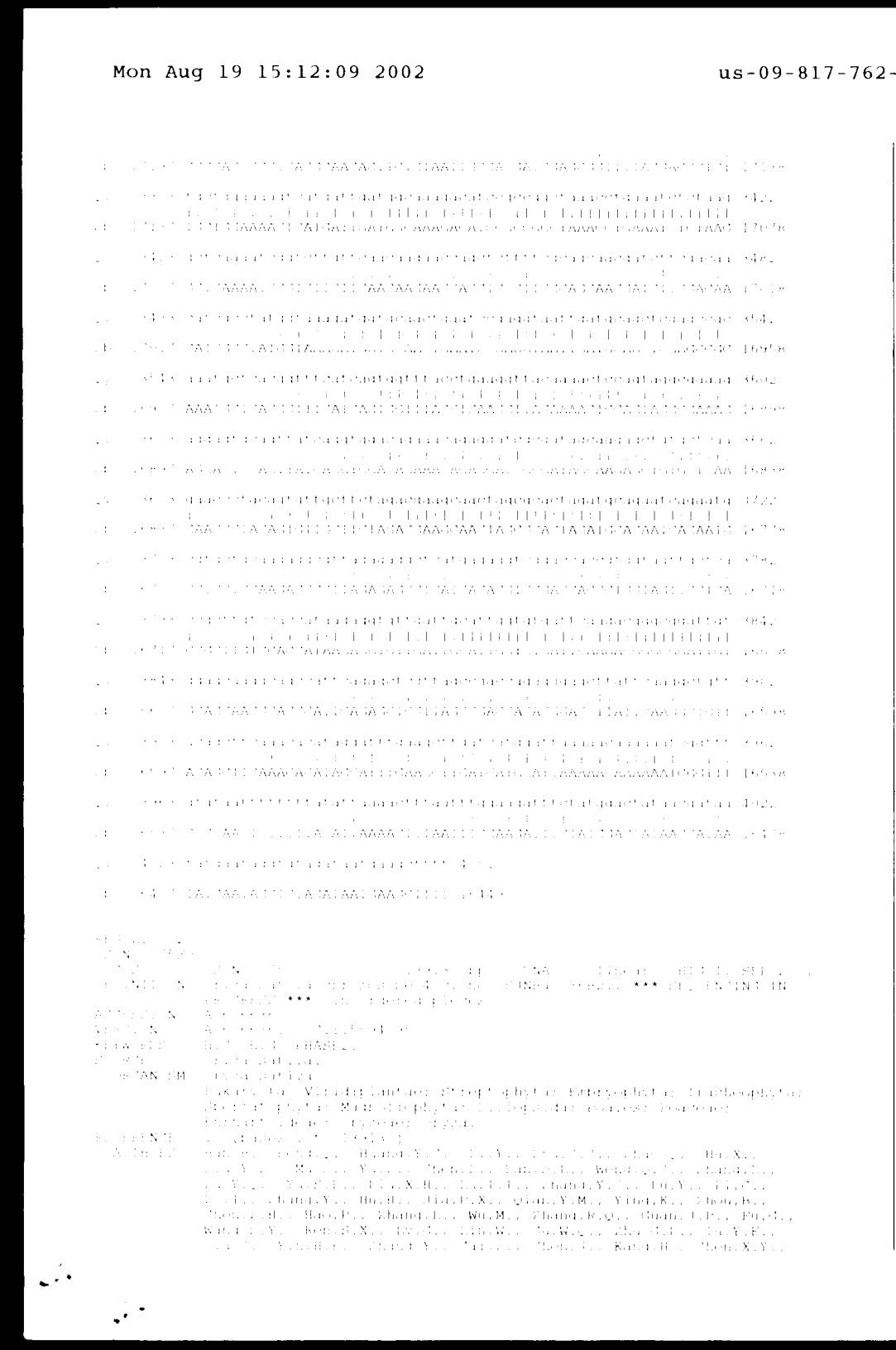


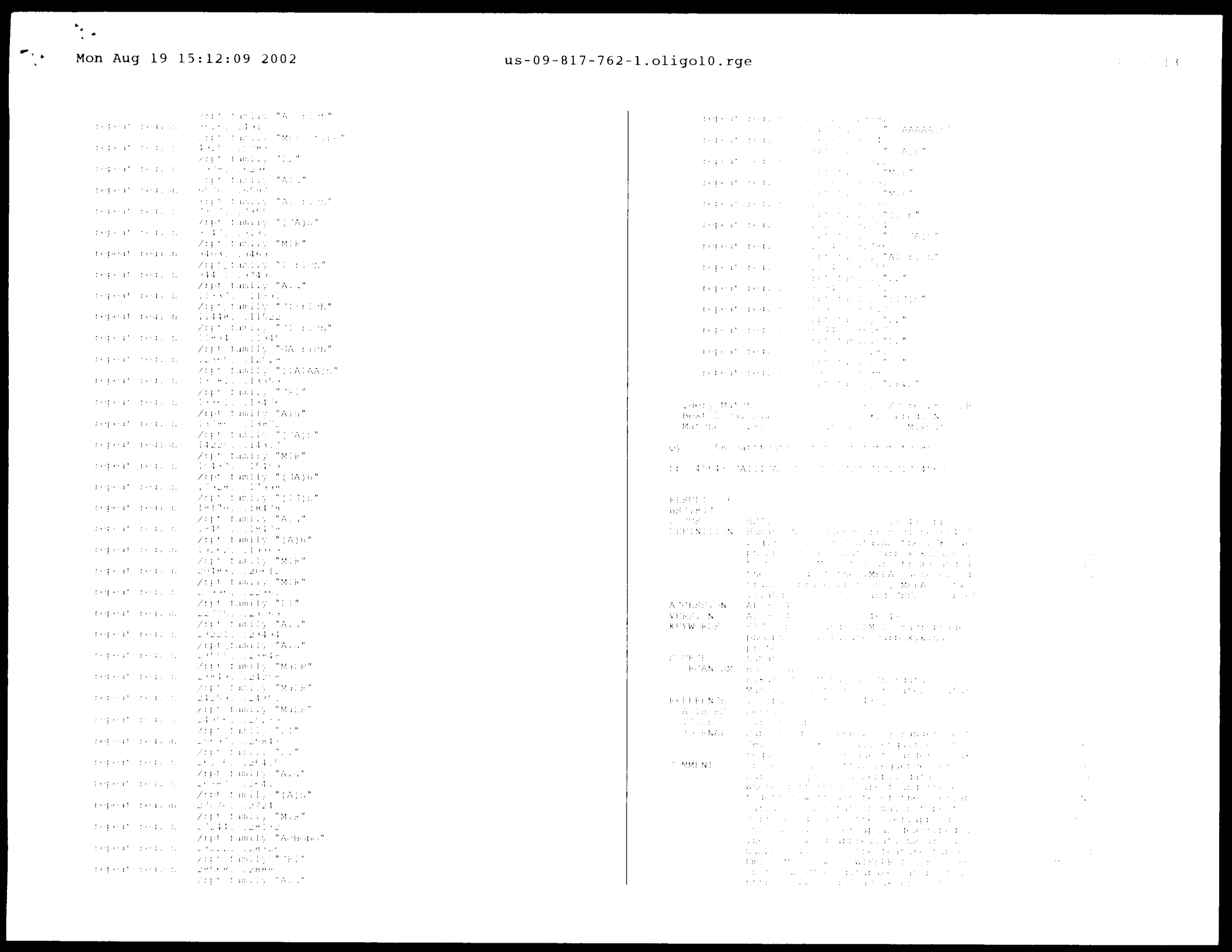


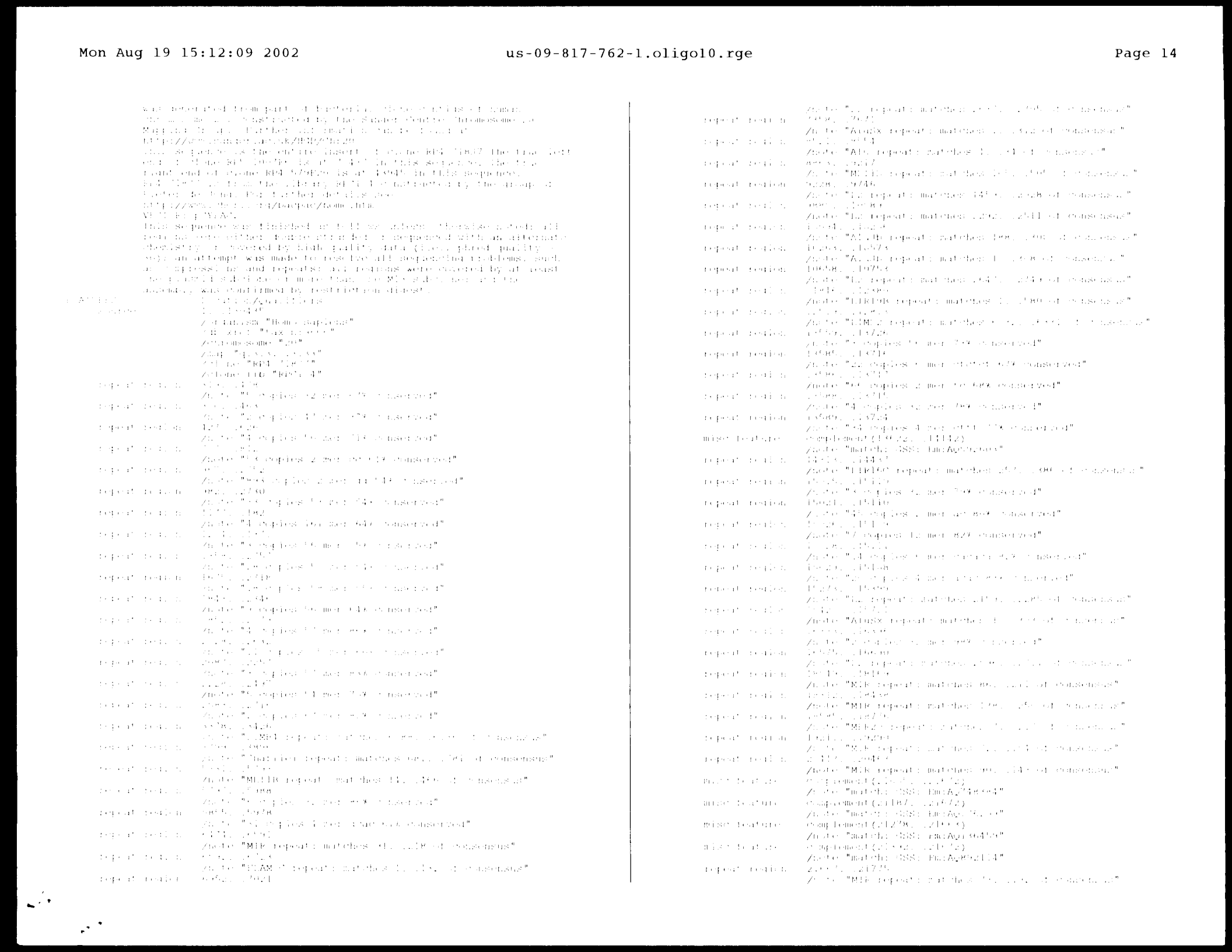


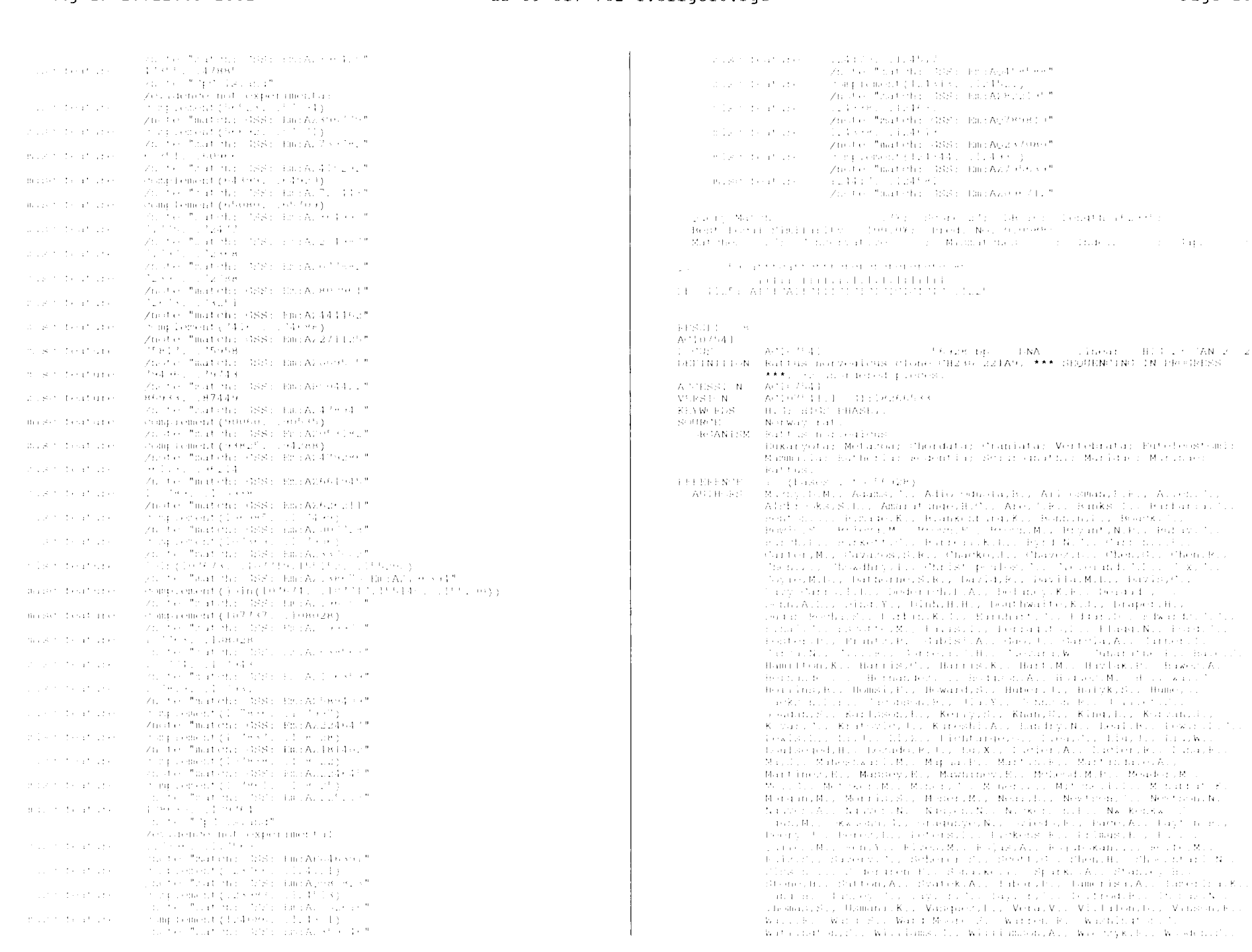


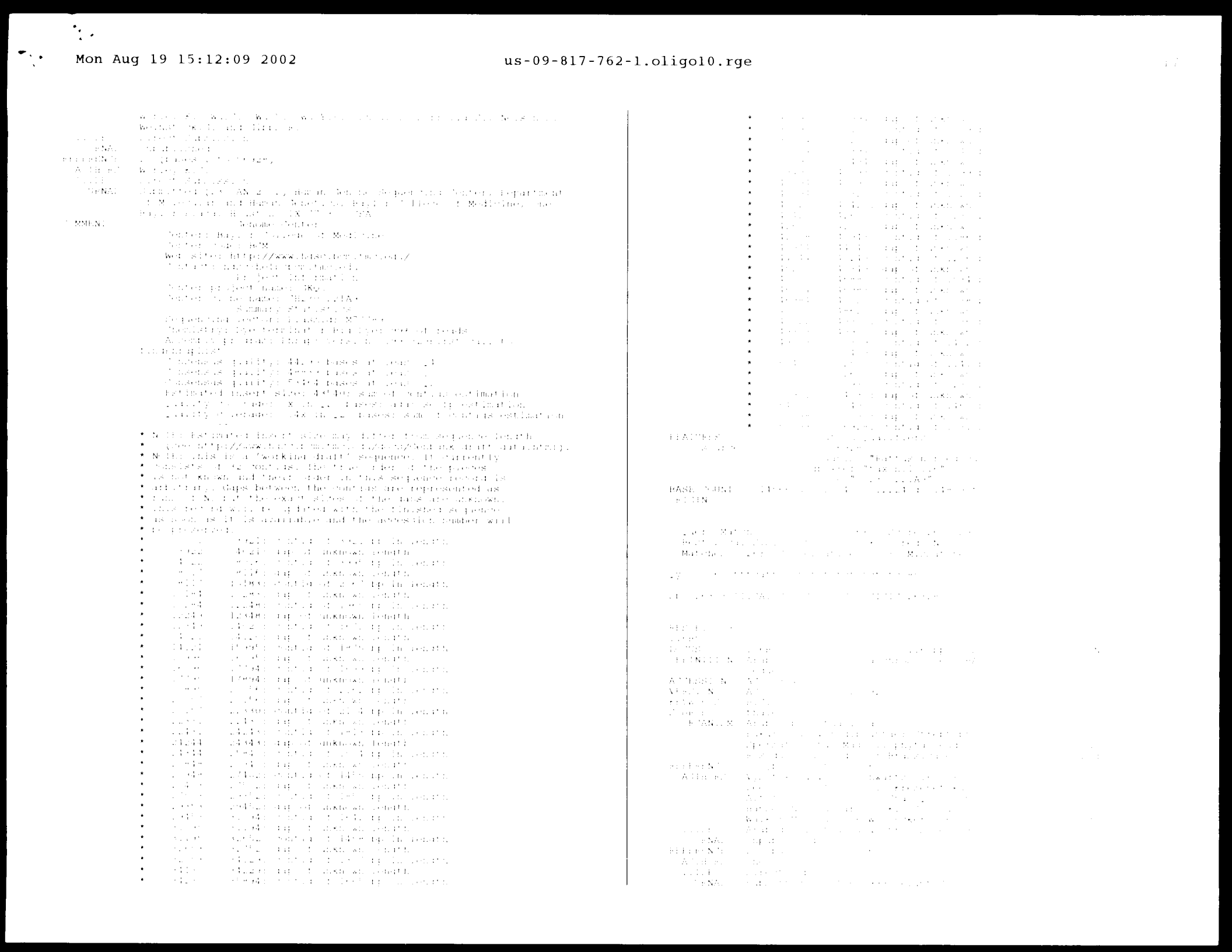


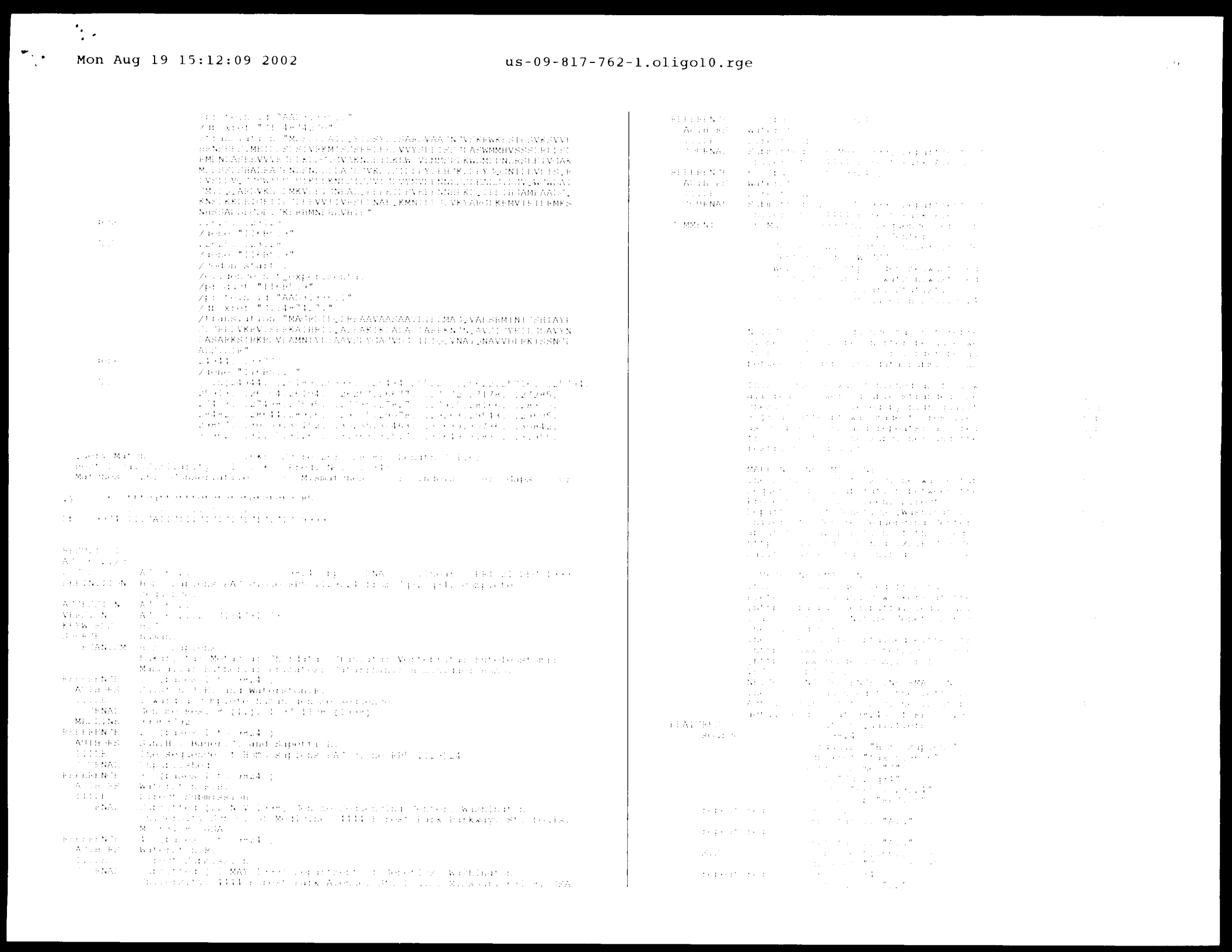


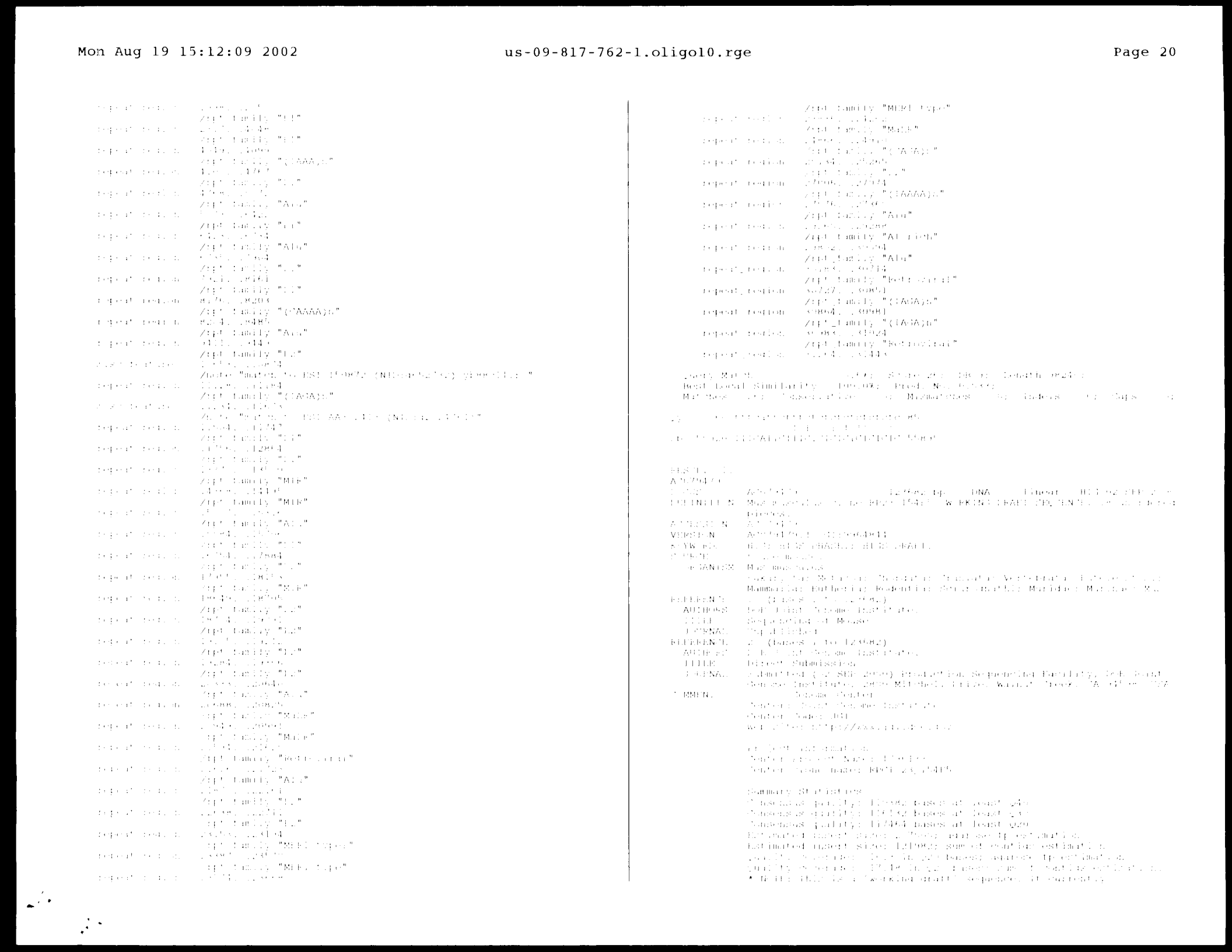


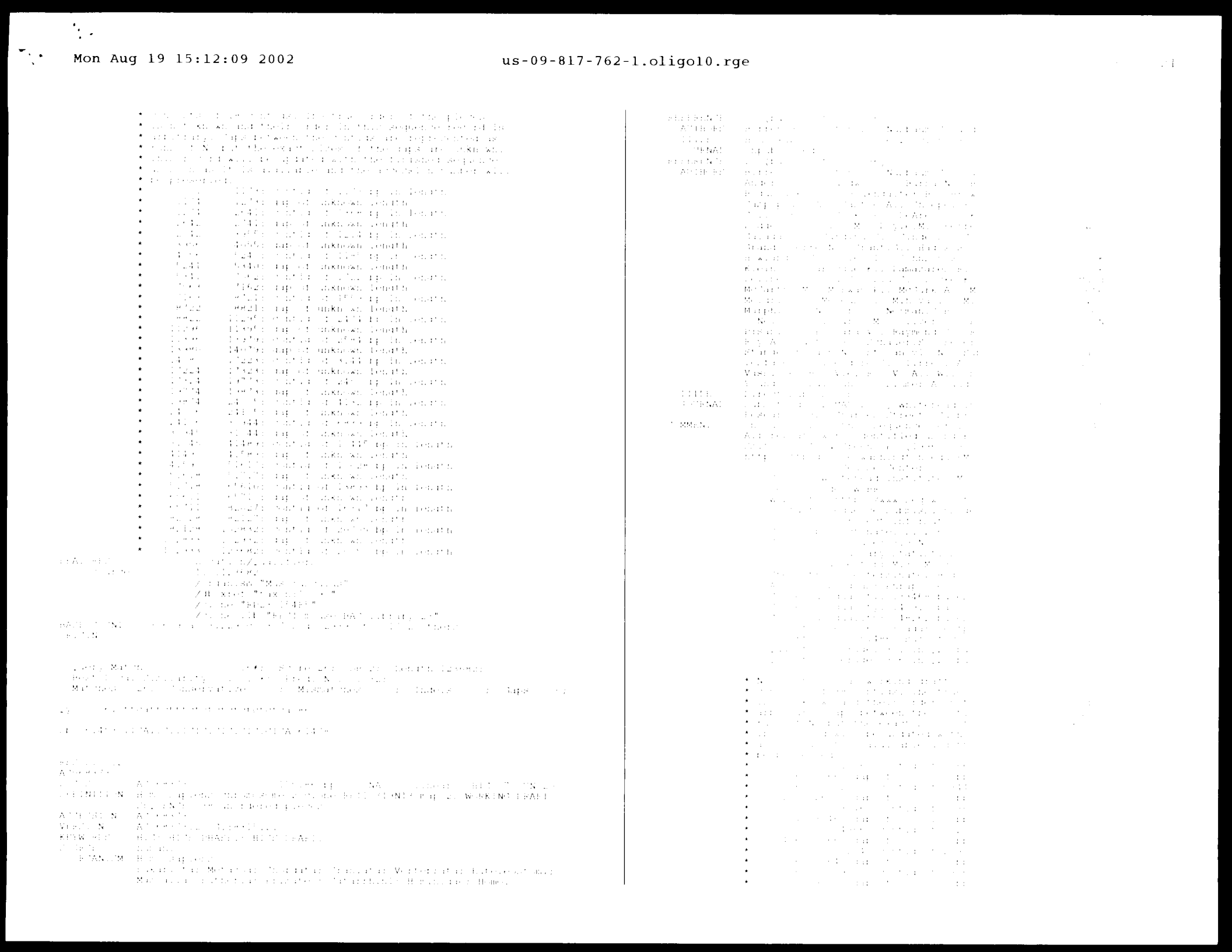


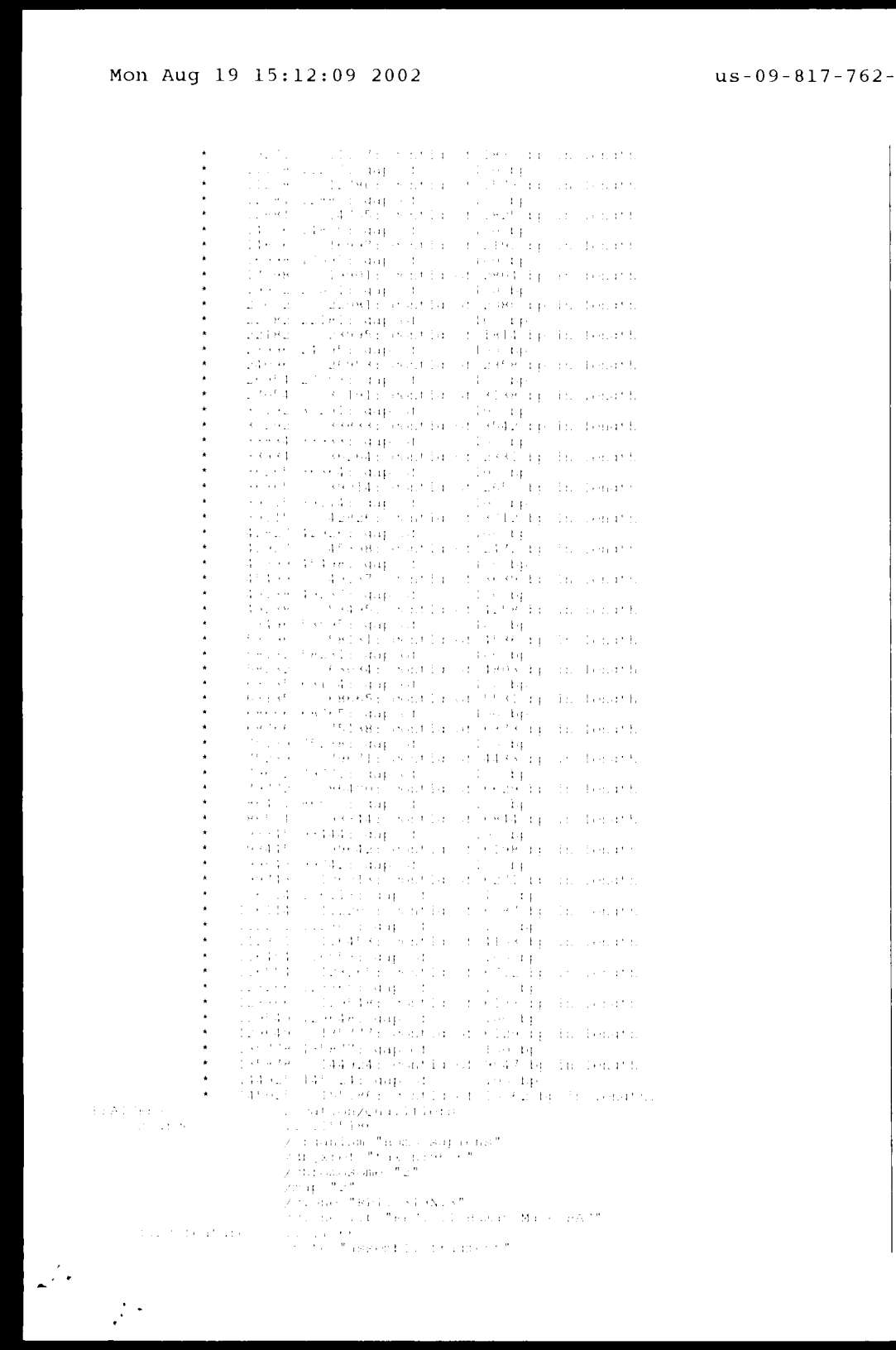


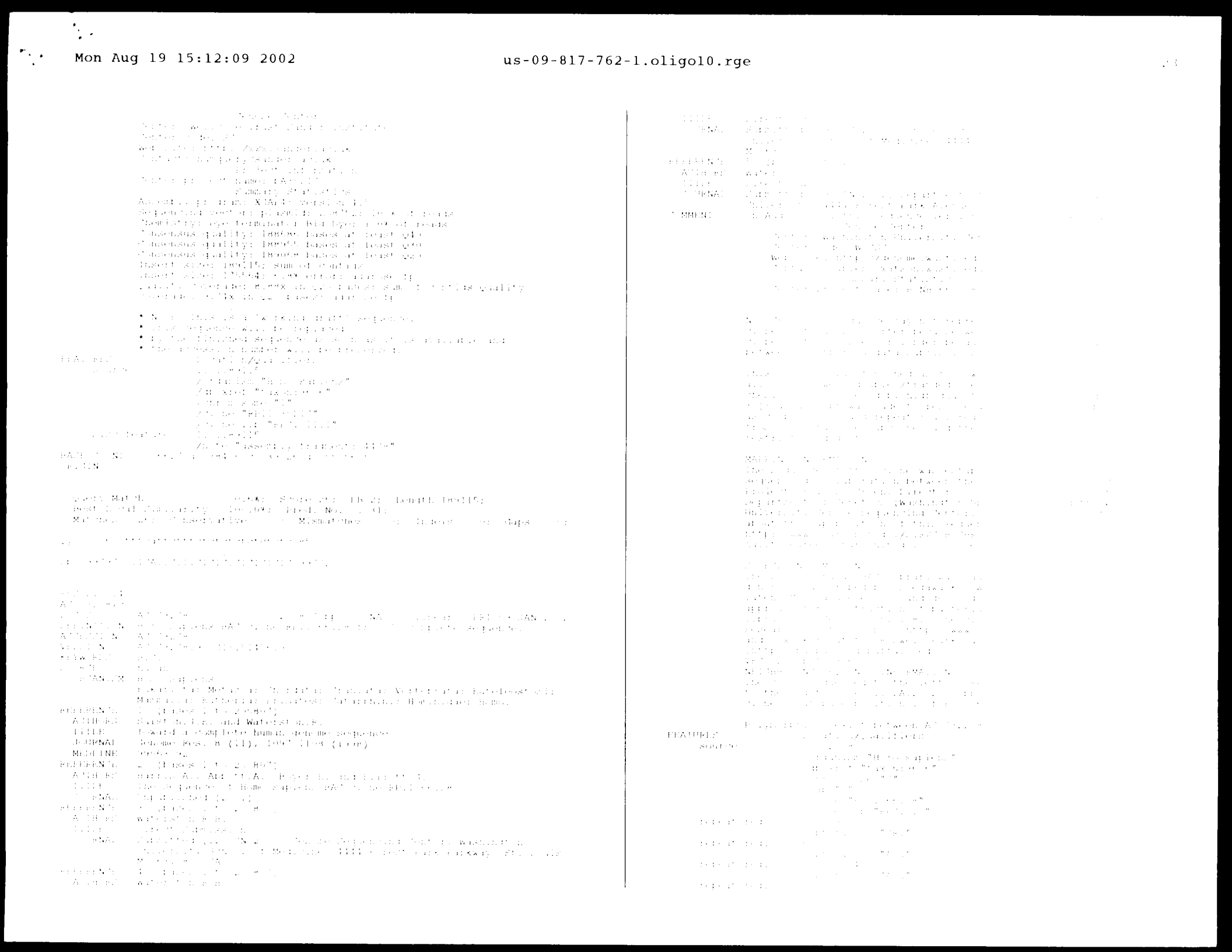












```
                     /rpt_family "Alu"
repeat_region        2245..2508
                     /rpt_family "A..."
repeat_region        2502..2828
                     /rpt_family "L2"
repeat_region        2811..2968
                     /rpt_family "C1 rich"
repeat_region        3055..3265
                     /rpt_family "Alu"
repeat_region        3238..3397
                     /rpt_family "MIR"
repeat_region        3387..3458
                     /rpt_family "C rich"
repeat_region        3456..3503
                     /rpt_family "(TTTTC)n"
repeat_region        3480..3784
                     /rpt_family "Alu"
repeat_region        3982..4295
                     /rpt_family "Alu"
repeat_region        4302..4477
                     /rpt_family "L1"
repeat_region        4528..4792
                     /rpt_family "Alu"
repeat_region        4791..4812
                     /rpt_family "(CA)n"
repeat_region        4852..4934
                     /rpt_family "L1"
repeat_region        4927..5144
                     /rpt_family "L1"
repeat_region        5145..5427
                     /rpt_family "Alu"
repeat_region        5428..5796
                     /rpt_family "L2"
repeat_region        5882..5915
                     /rpt_family "Alu"
repeat_region        5961..6094
                     /rpt_family "L1"
repeat_region        6084..6430
                     /rpt_family "(TG)n"
repeat_region        6430..6681
                     /rpt_family "Alu"
repeat_region        6705..6825
                     /rpt_family "(CA)n"
repeat_region        6826..6942
                     /rpt_family "(TA)n"
repeat_region        6955..7226
                     /rpt_family "Alu"
repeat_region        7206..7315
                     /rpt_family "AT_rich"
repeat_region        7228..7369
                     /rpt_family "Alu"
repeat_region        7351..7564
                     /rpt_family "AT rich"
repeat_region        7574..7756
                     /rpt_family "AT rich"
repeat_region        7808..8008
                     /rpt_family "Alu"
repeat_region        7986..8017
                     /rpt_family "AT rich"
repeat_region        8151..8258
                     /rpt_family "TCC60"
repeat_region        8254..8438
                     /rpt_family "MIRE"
repeat_region        8416..8543
                     /rpt_family "TCC60"
repeat_region        8571..8799
                     /rpt_family "MIR"
repeat_region        8750..8952
                     /rpt_family "Alu"
repeat_region        8982..9262
                     /rpt_family "Alu"
repeat_region        9256..9422
                     /rpt_family "AT rich"
repeat_region        9472..9600
                     /rpt_family "Alu"
repeat_region        9590..9836
                     /rpt_family "(CA)n"
repeat_region        10027..10113
                     /rpt_family "A rich"
repeat_region        10012..10476
                     /rpt_family "T rich"
repeat_region        10451..10754
                     /rpt_family "Alu"
repeat_region        11141..11501
                     /rpt_family "L2"
repeat_region        11502..11745
                     /rpt_family "Alu"
repeat_region        11527..11600
                     /rpt_family "AT_rich"
repeat_region        11746..11768
                     /rpt_family "AT_rich"
repeat_region        12550..12690
                     /rpt_family "Alu"
repeat_region        13008..13140
                     /rpt_family "L1"
repeat_region        13341..13440
                     /rpt_family "Alu"
repeat_region        13441..13692
                     /rpt_family "L1"
repeat_region        13693..13978
                     /rpt_family "Alu"

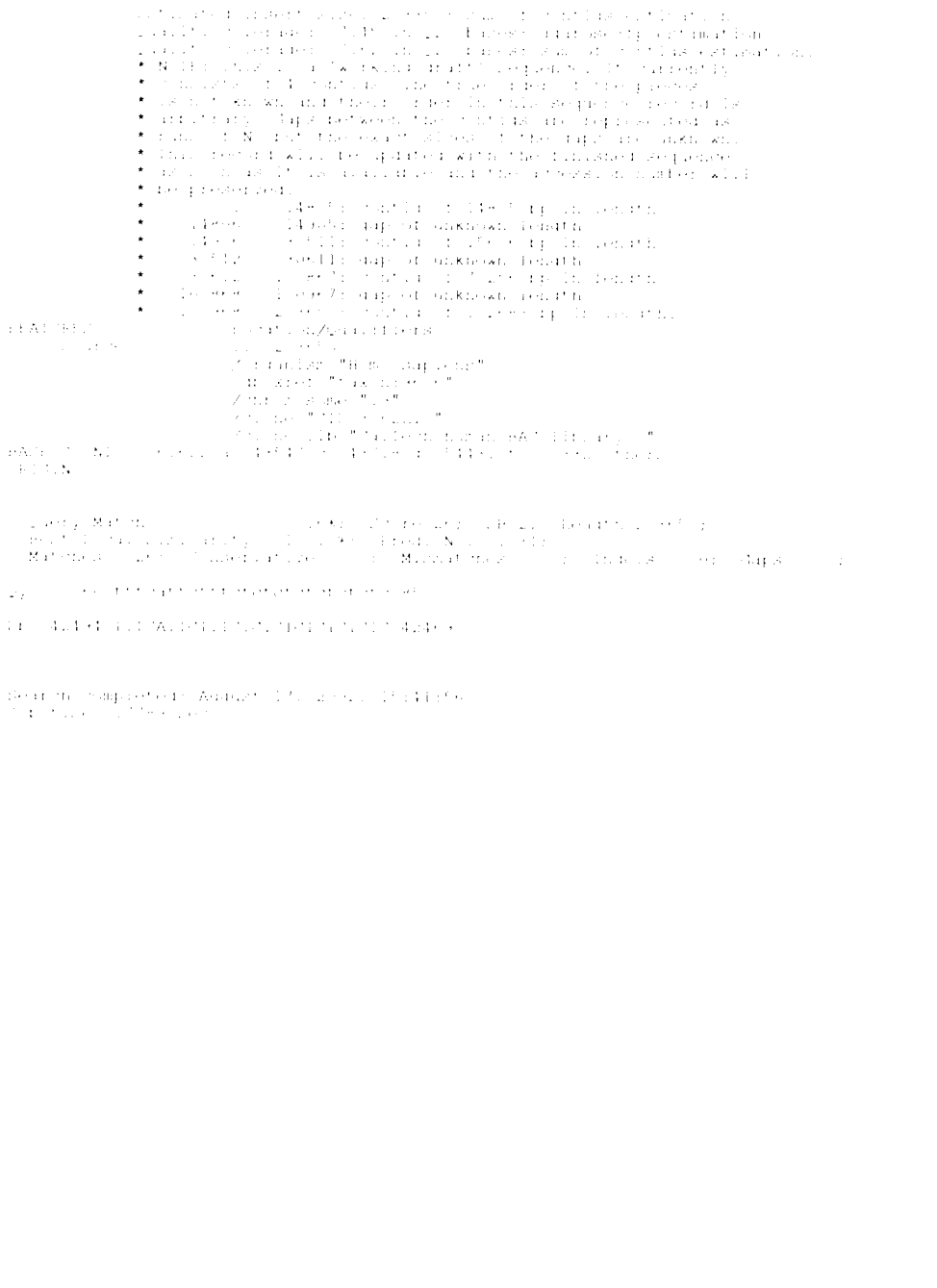
  Query Match              0.6%;  Score 26;  DB 9;  Length 200807;
  Best Local Similarity   100.0%;  Pred. No. 0.041;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

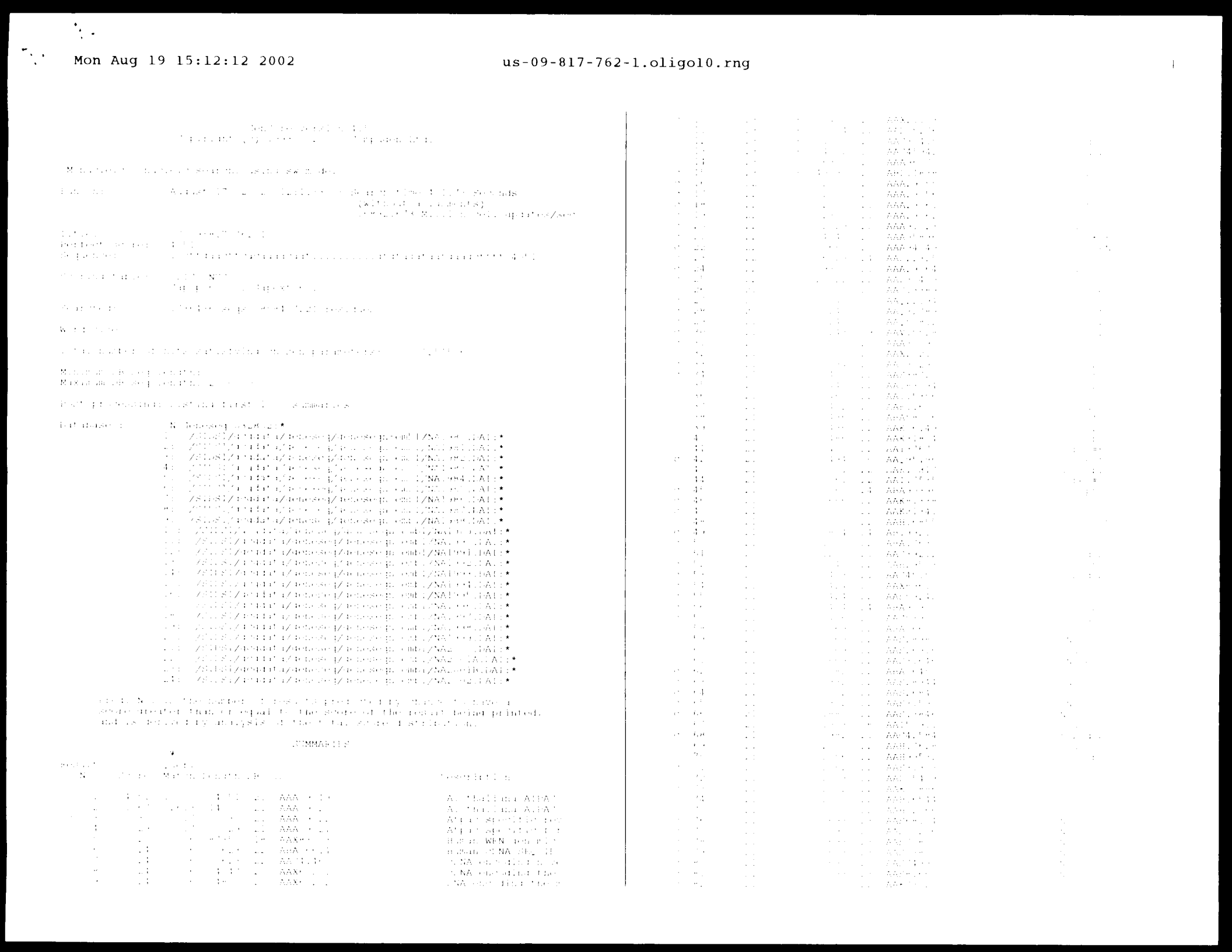
Qy      60 tttcattctttctctctctctctc 85
           ||||||||||||| |||||||||||
Db 197488 tttcattctttctctctctctctc 197438

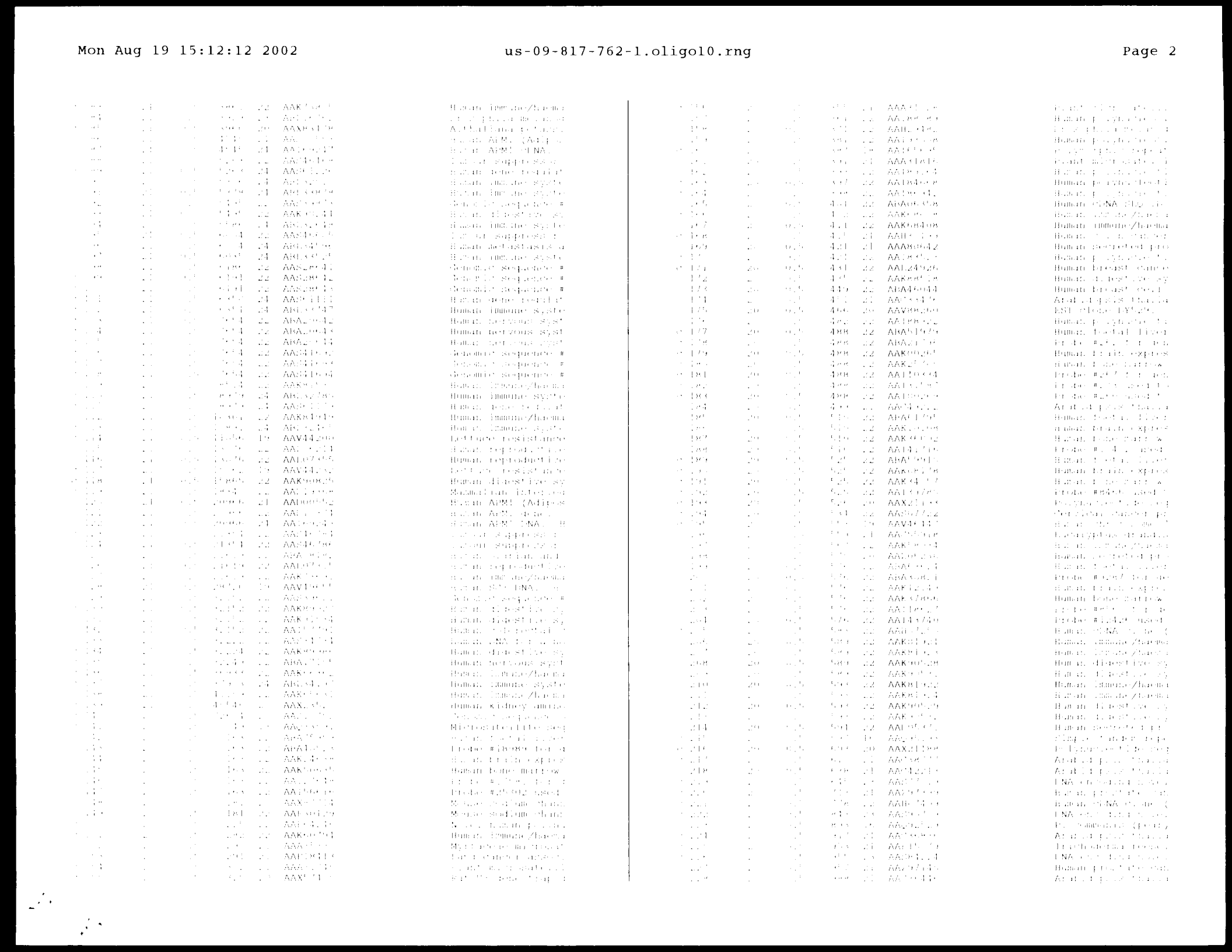

RESULT 15
AC092296/c
LOCUS       AC092296               203650 bp    DNA     linear   HTG 05-JUL-2001
DEFINITION  Homo sapiens chromosome 19 clone CTD-2161I1, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC092296
VERSION     AC092296.1  GI:14589485
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203650)
  AUTHORS   Jek Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203650)
  AUTHORS   Jek Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 JUL 2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT             Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

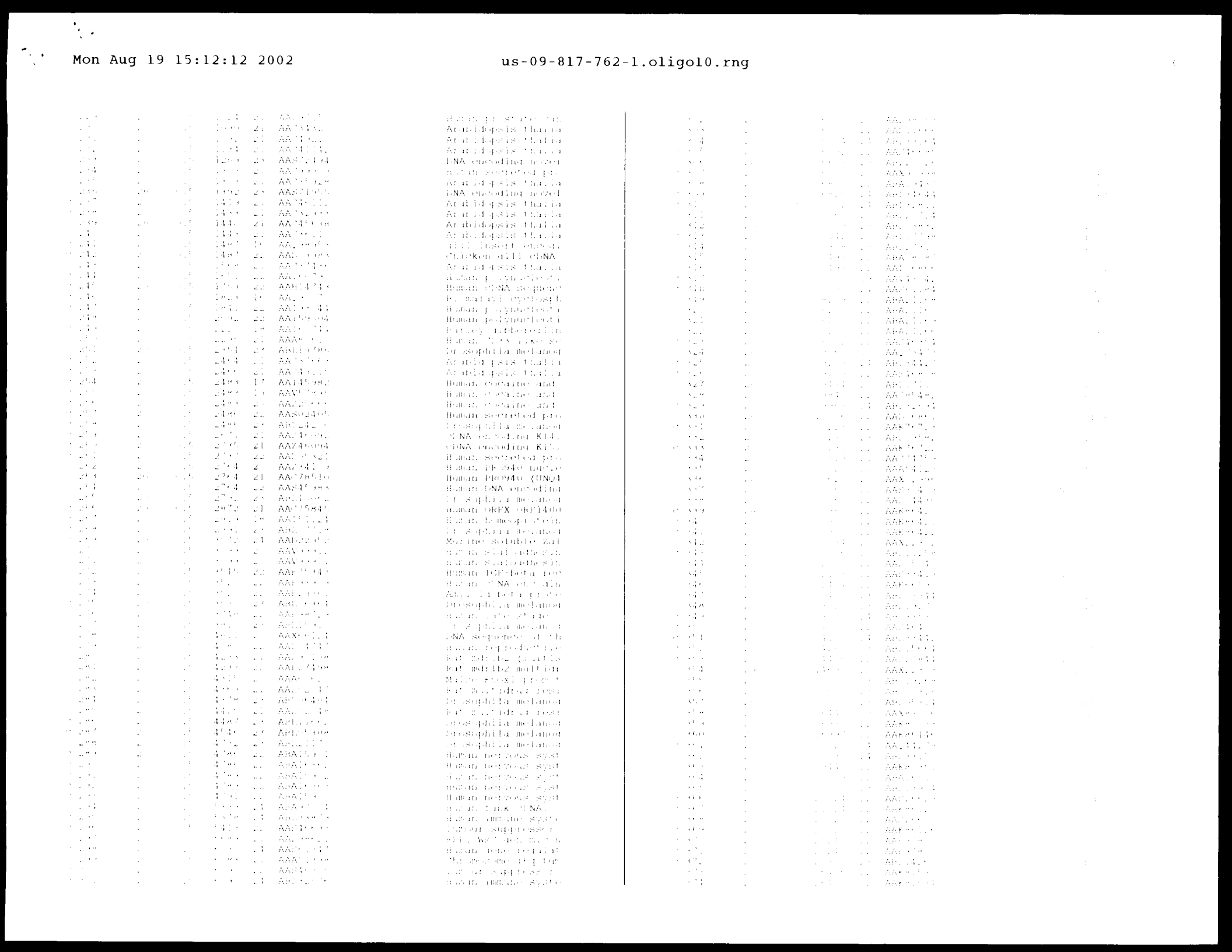
            Project Information
            Center Project Name: Gib2161_Jek0303
            Center clone name: CTD-2161I1

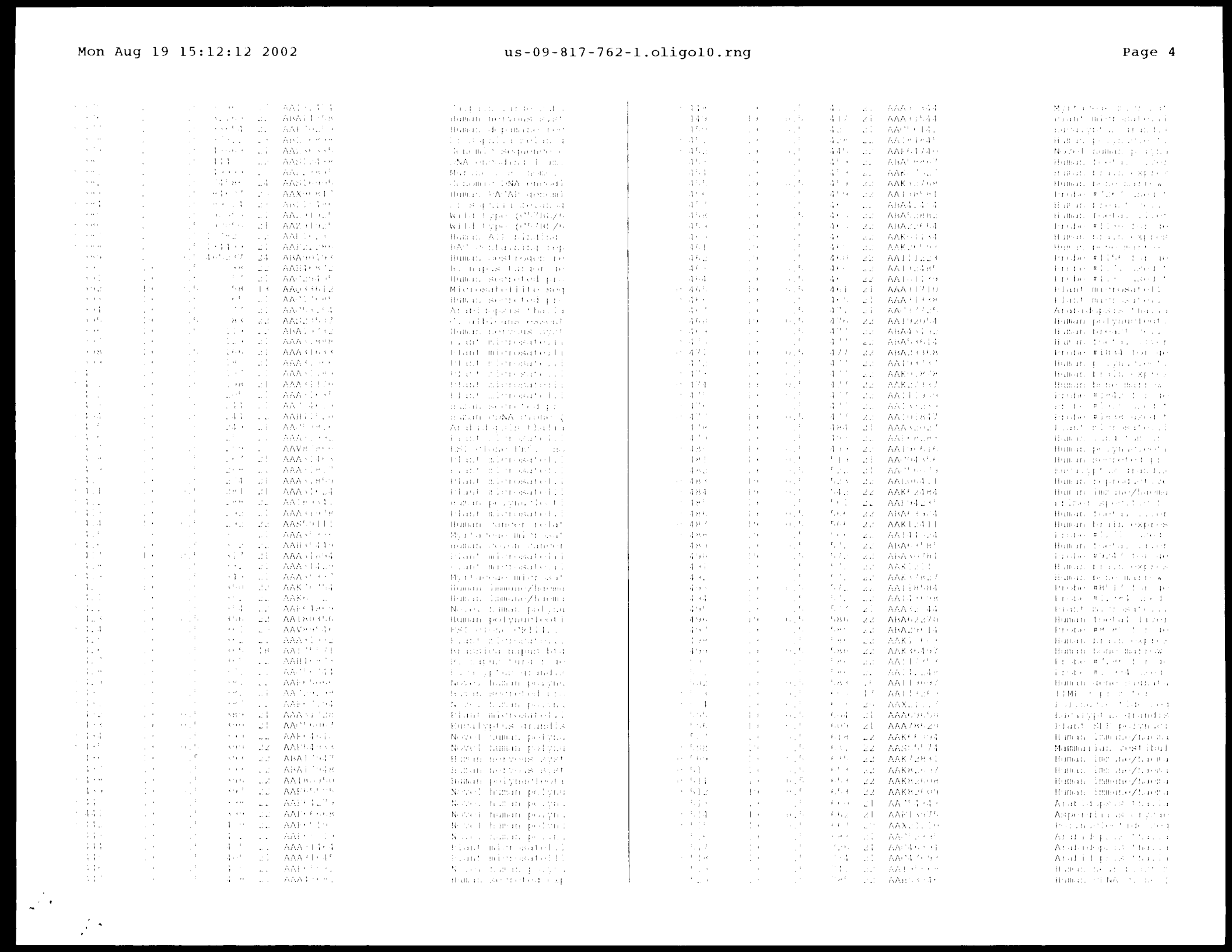
            Summary Statistics
            Consensus quality: 197185 bases at least q40
            Consensus quality: 202398 bases at least q20
            Consensus quality: 202821 bases at least q20
            Estimated insert size: 207960; agarose fp estimation
```

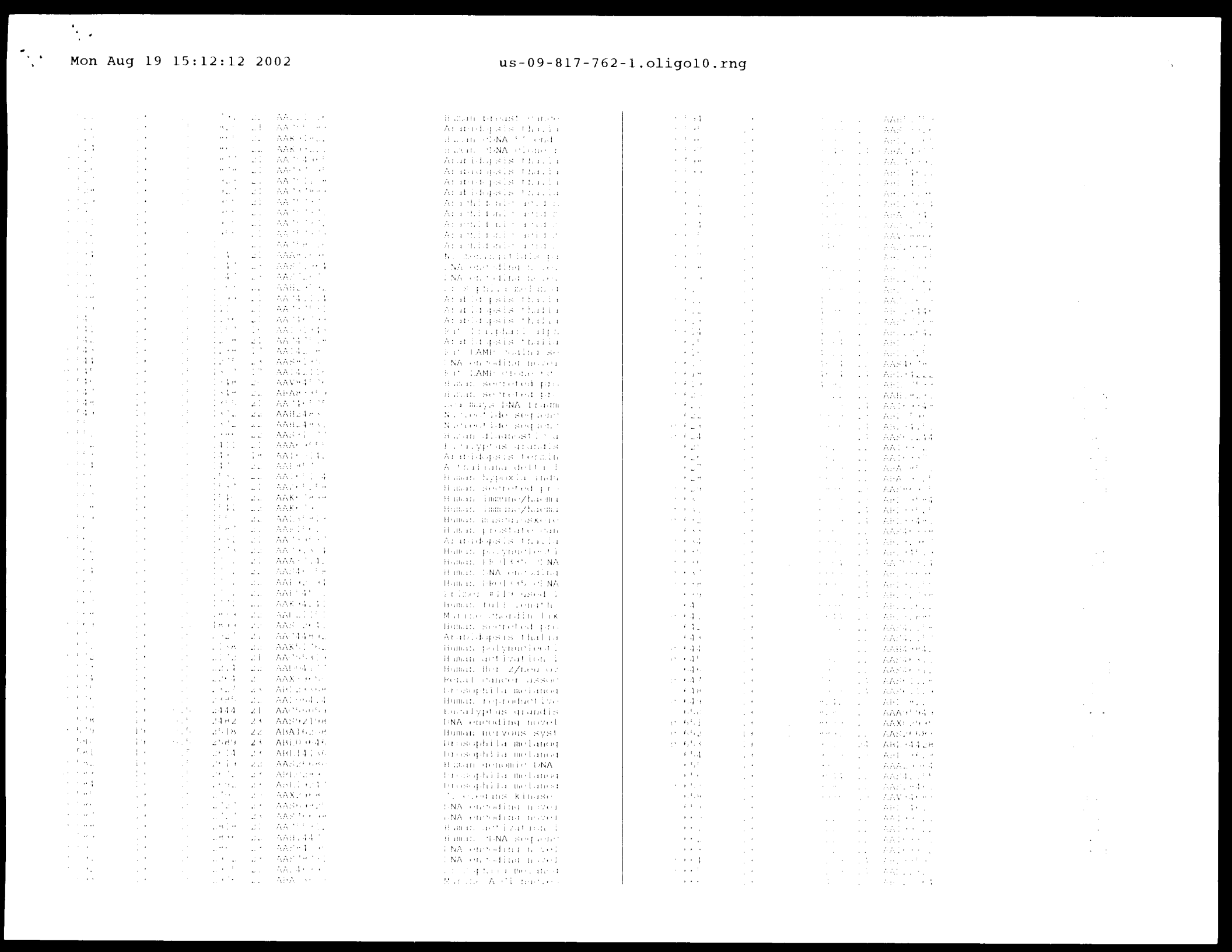


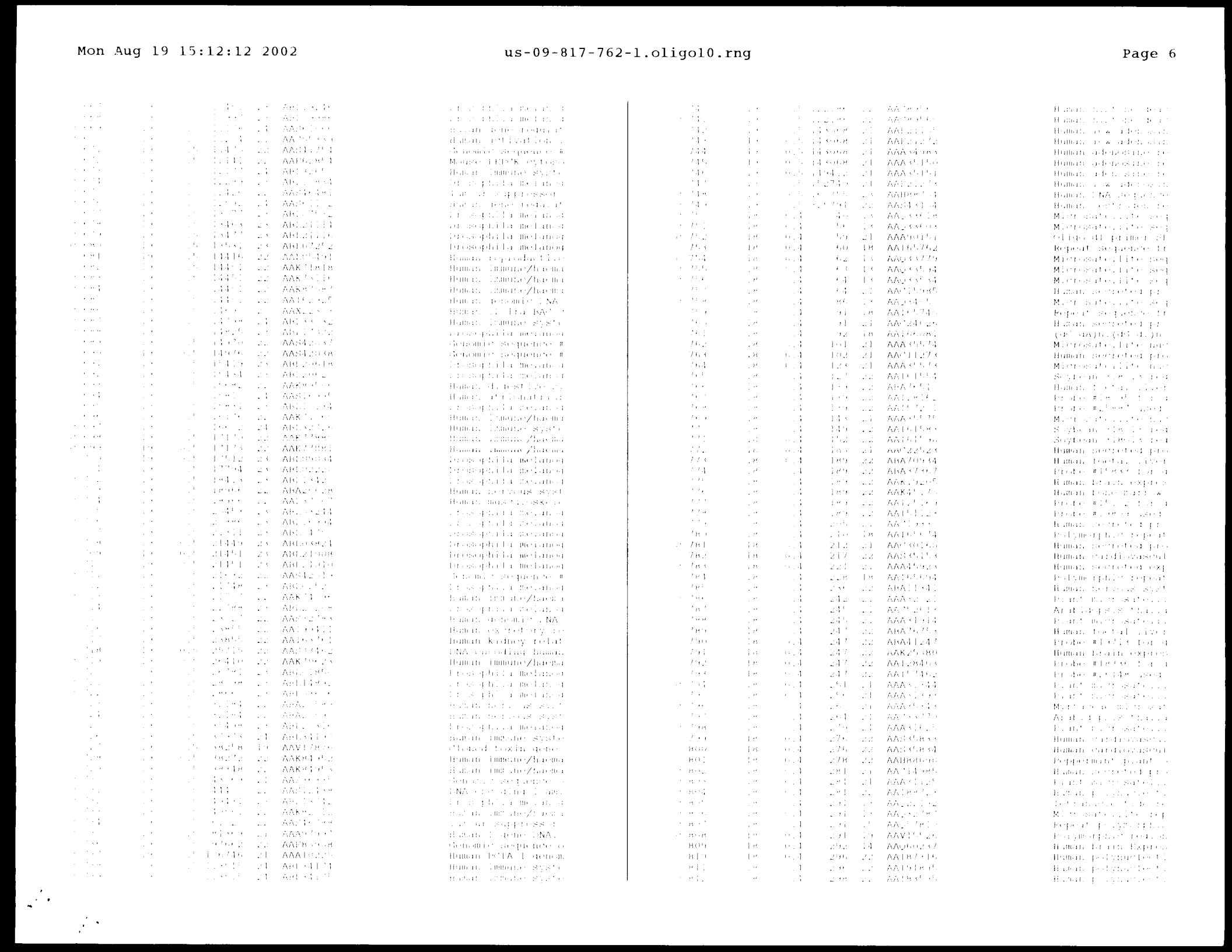


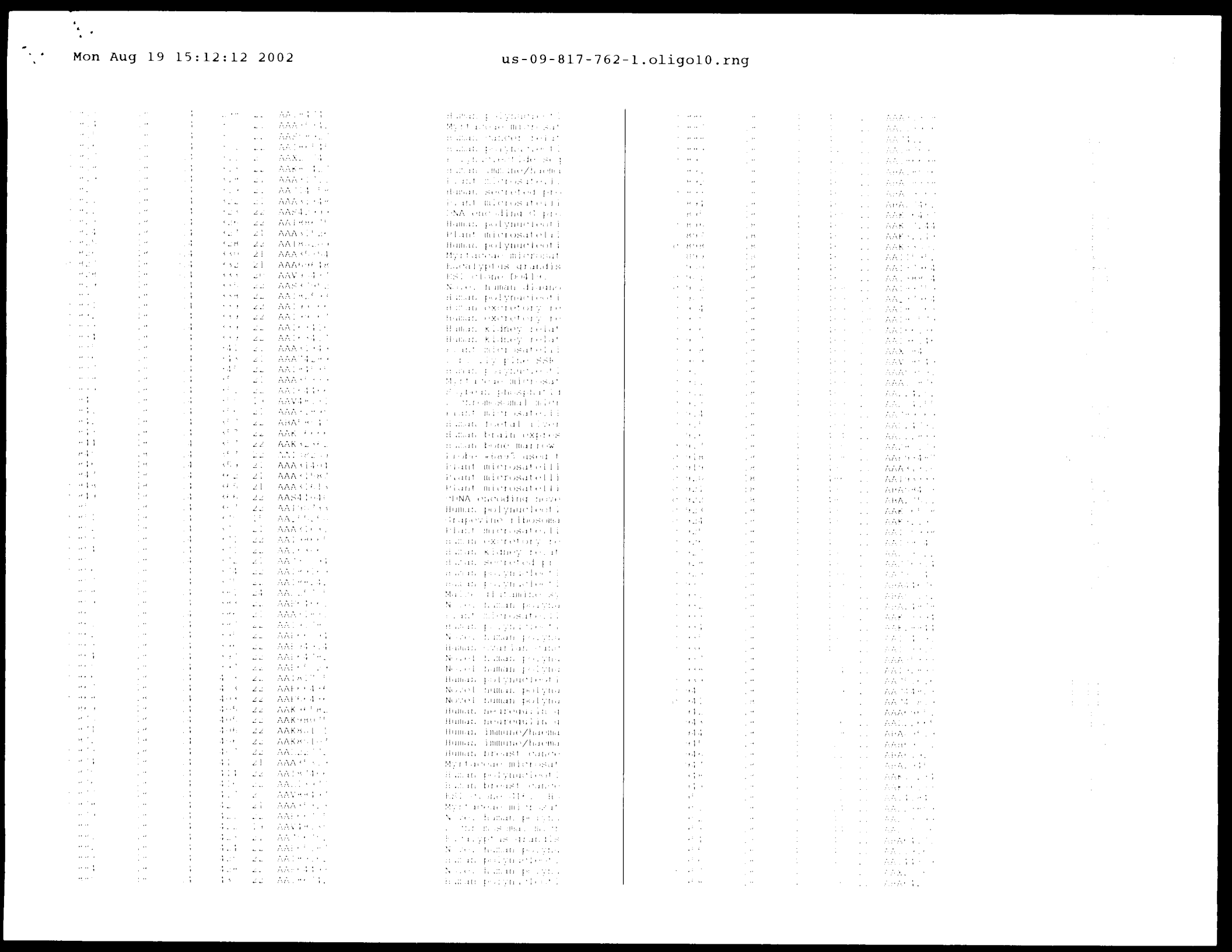


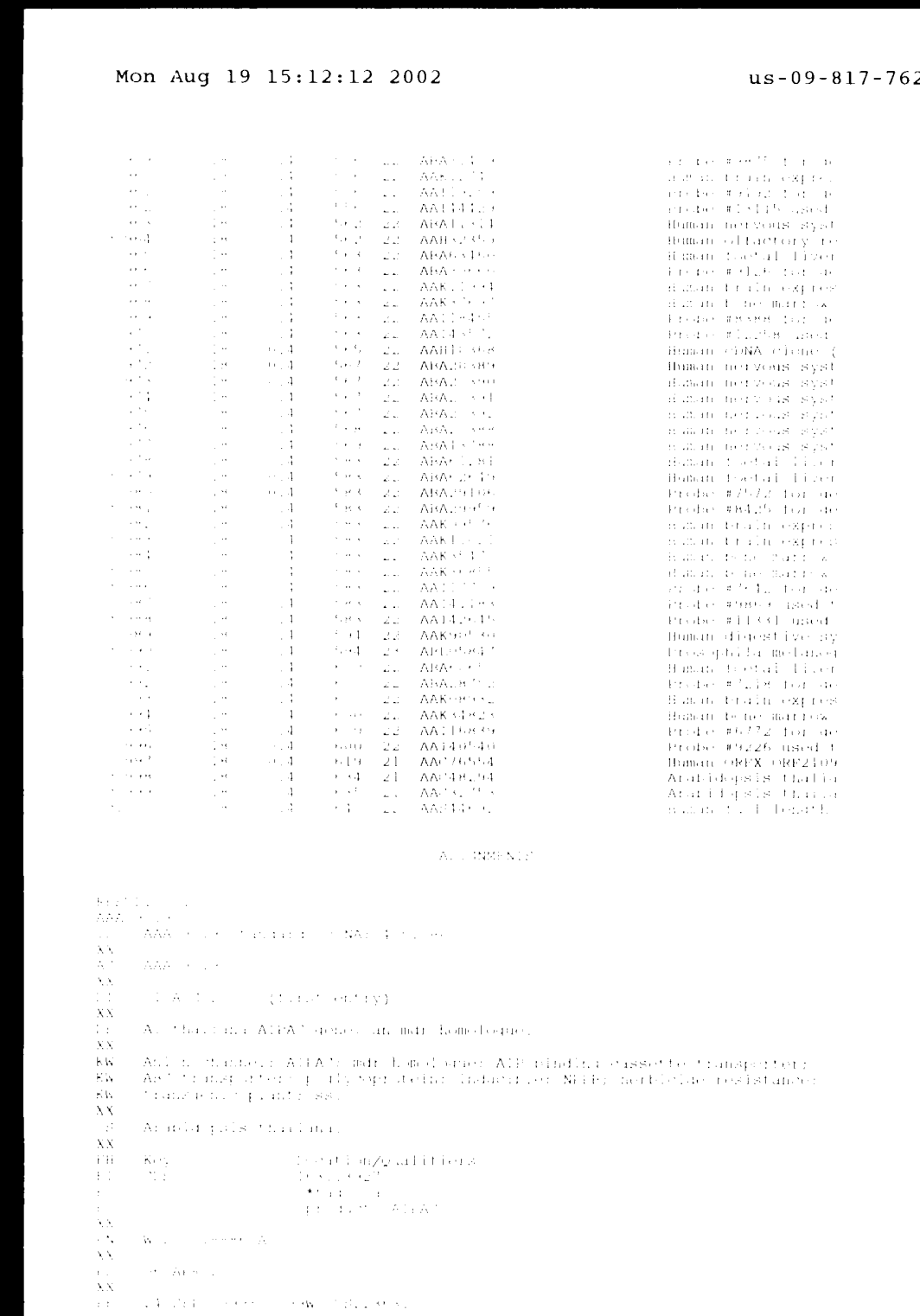


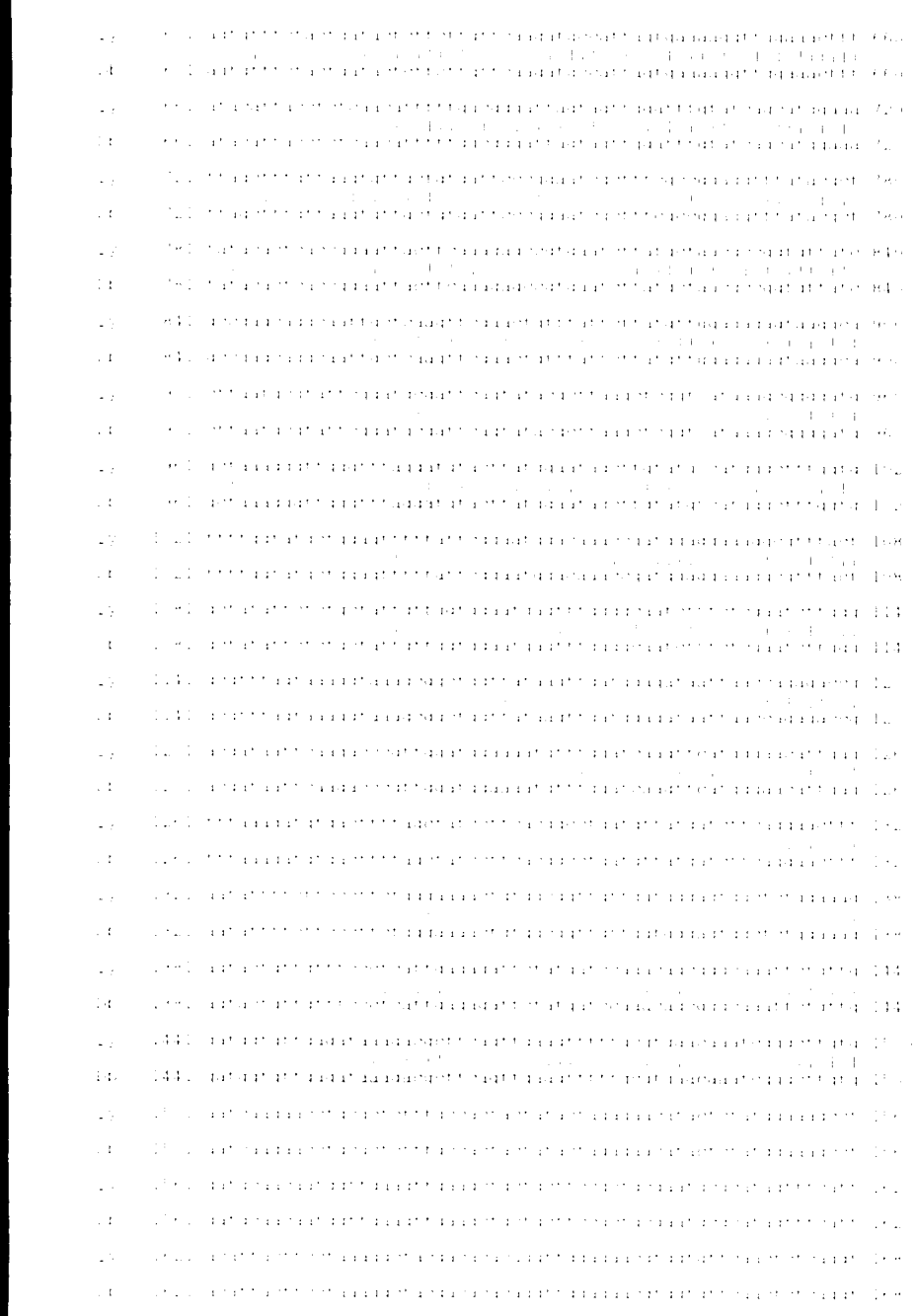


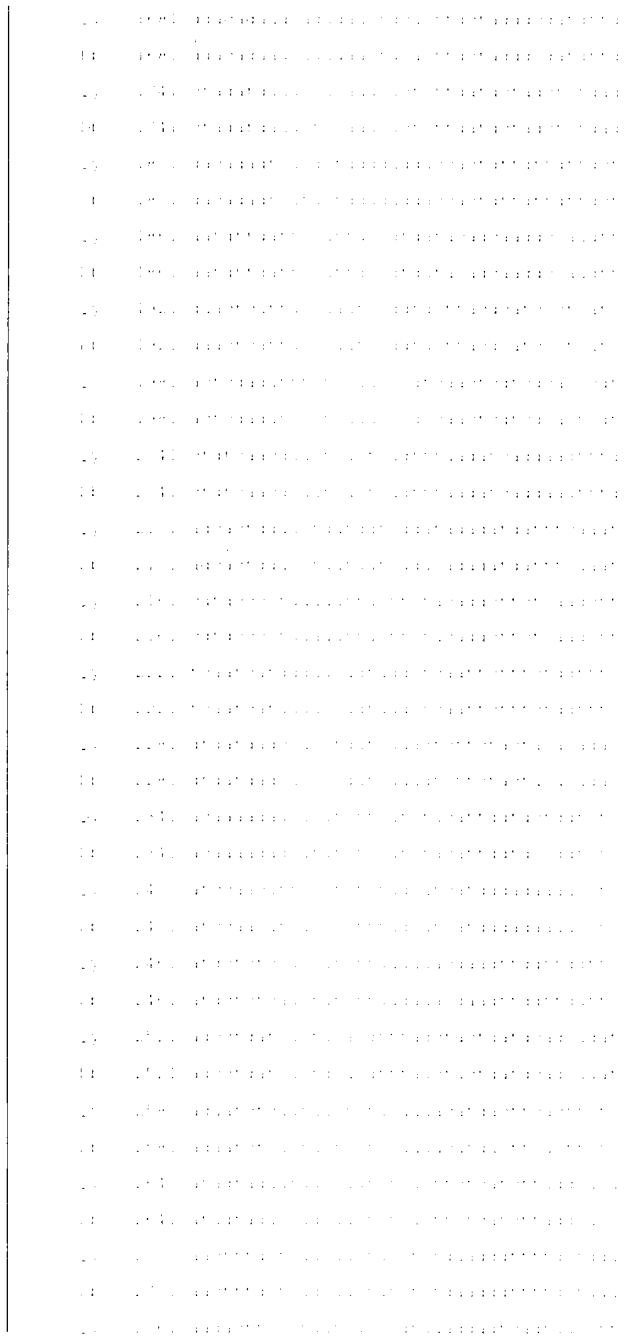


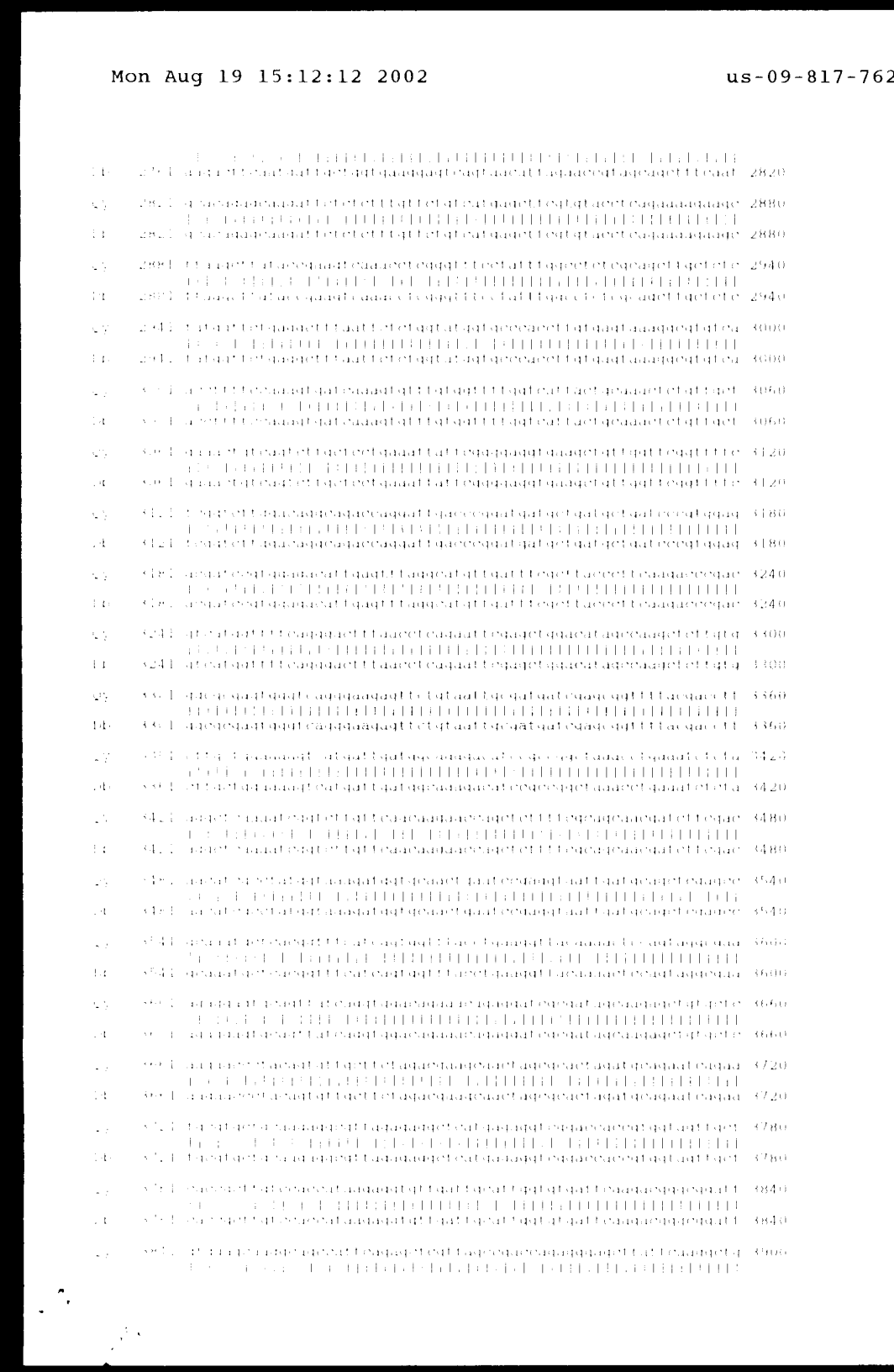




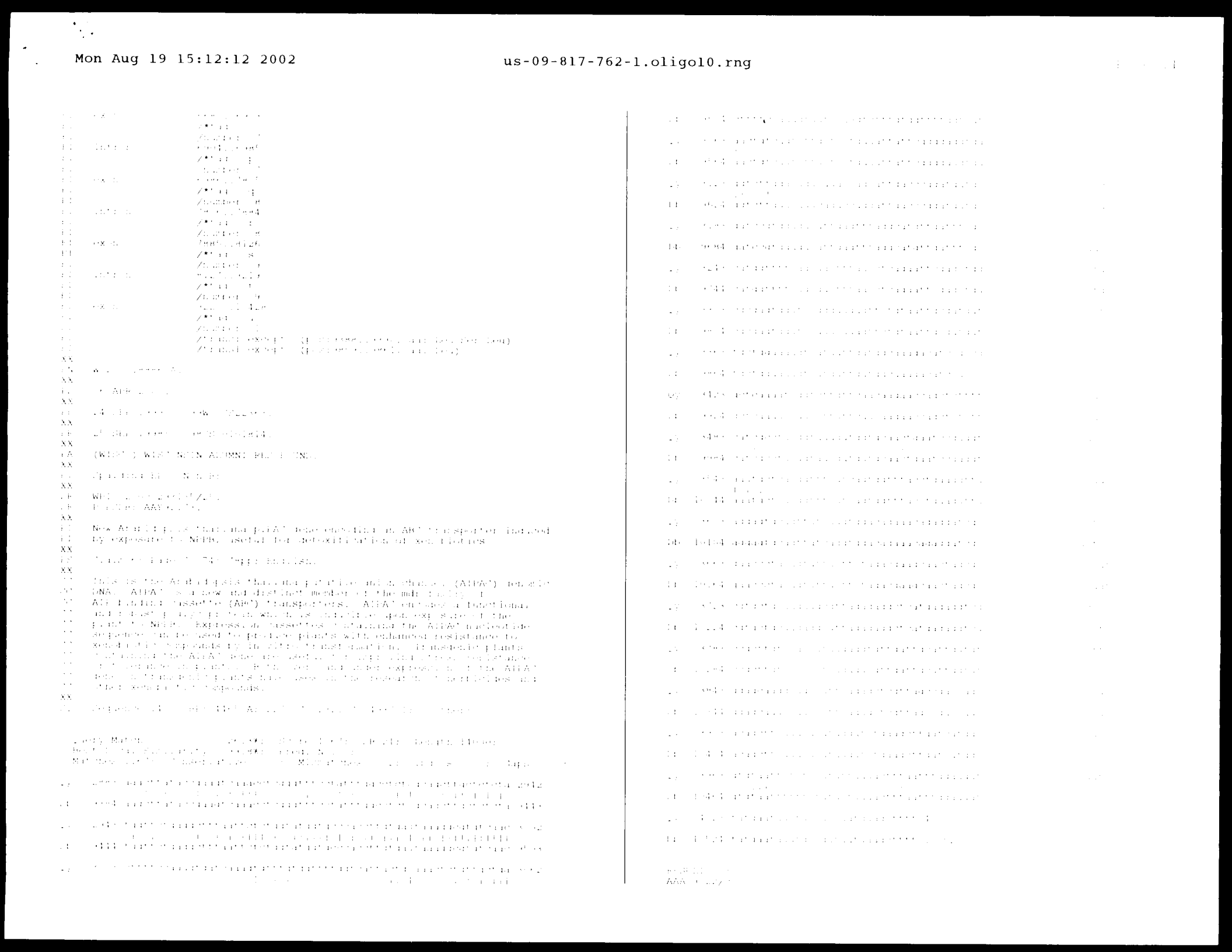


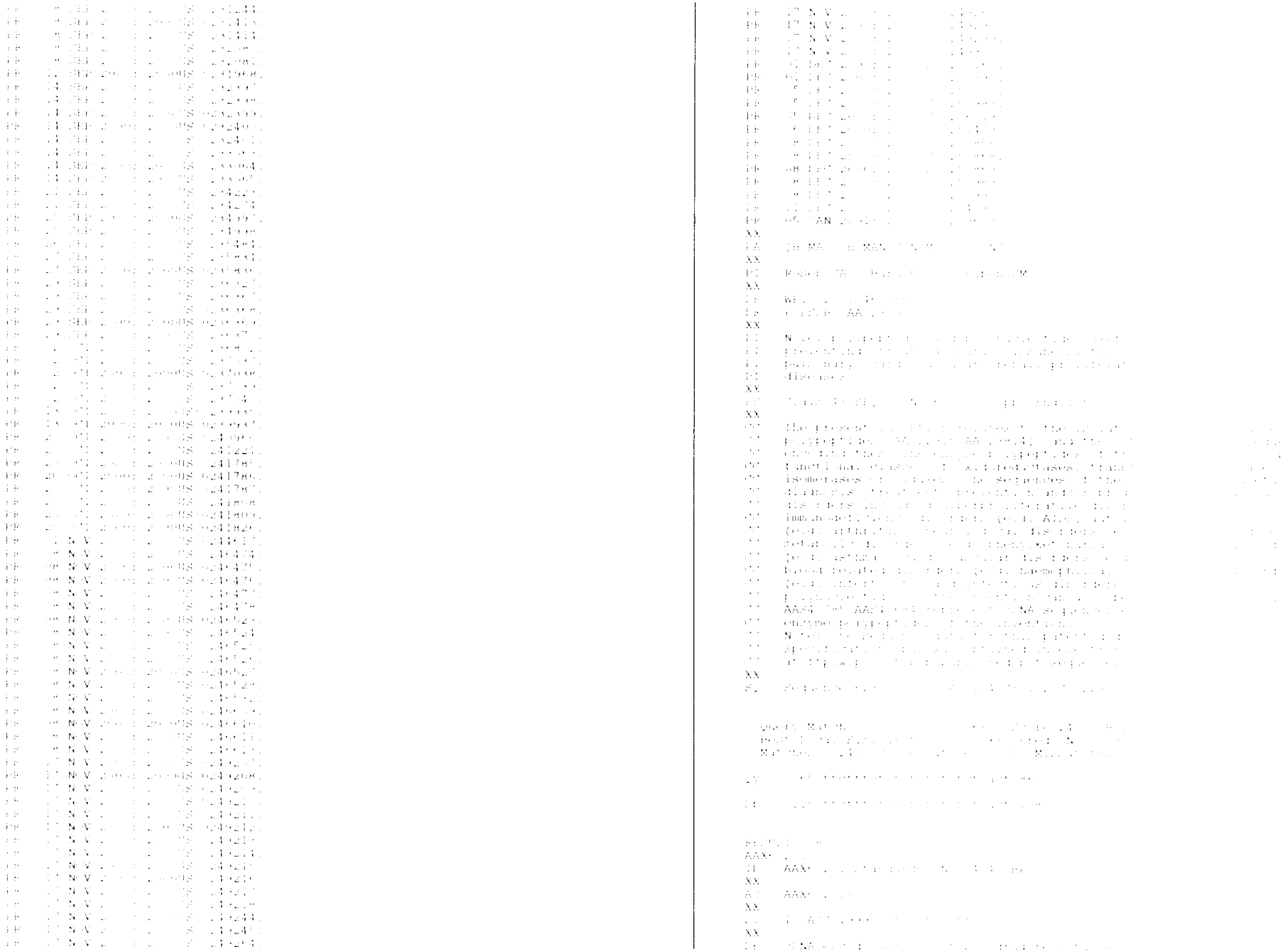


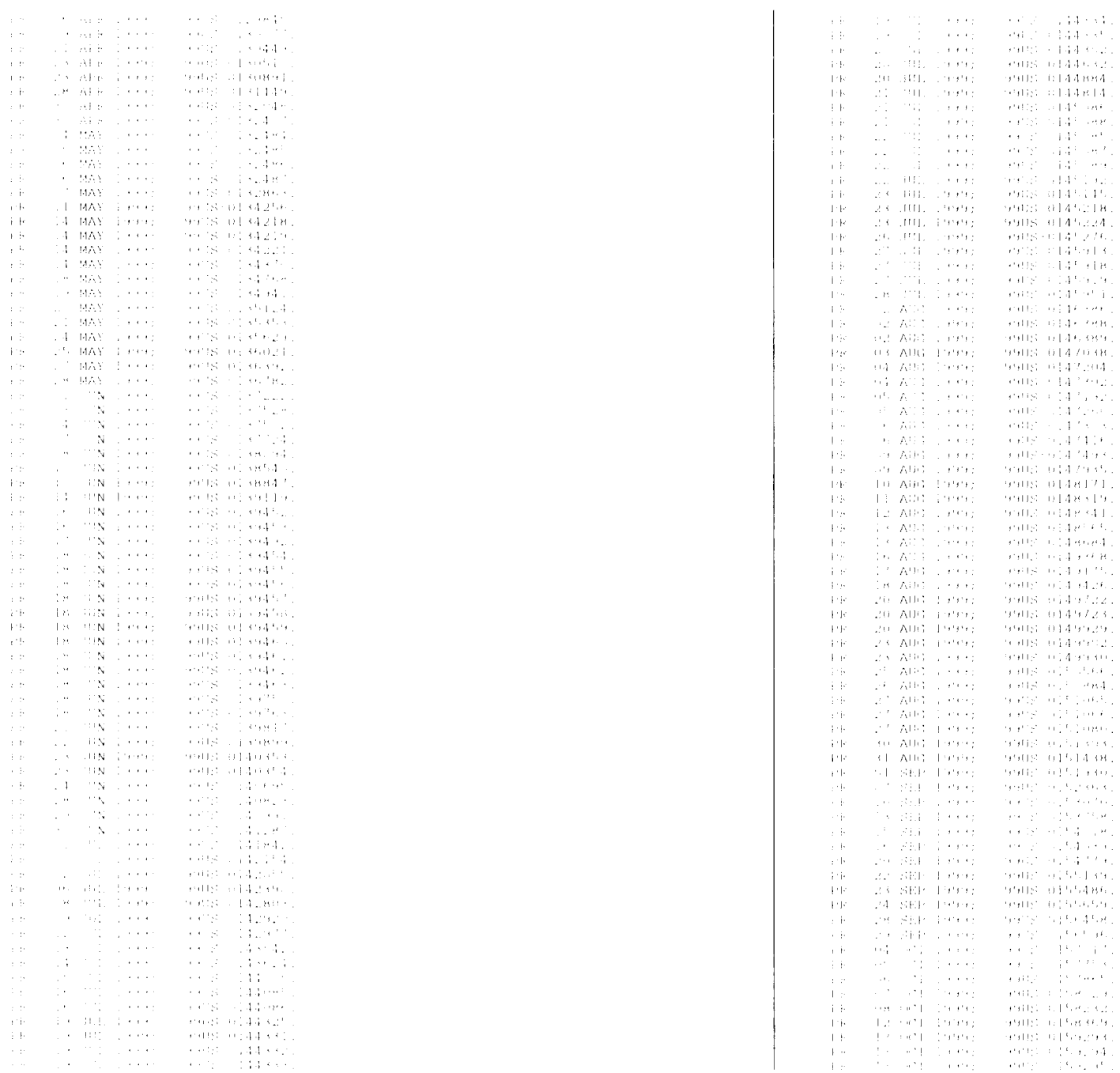


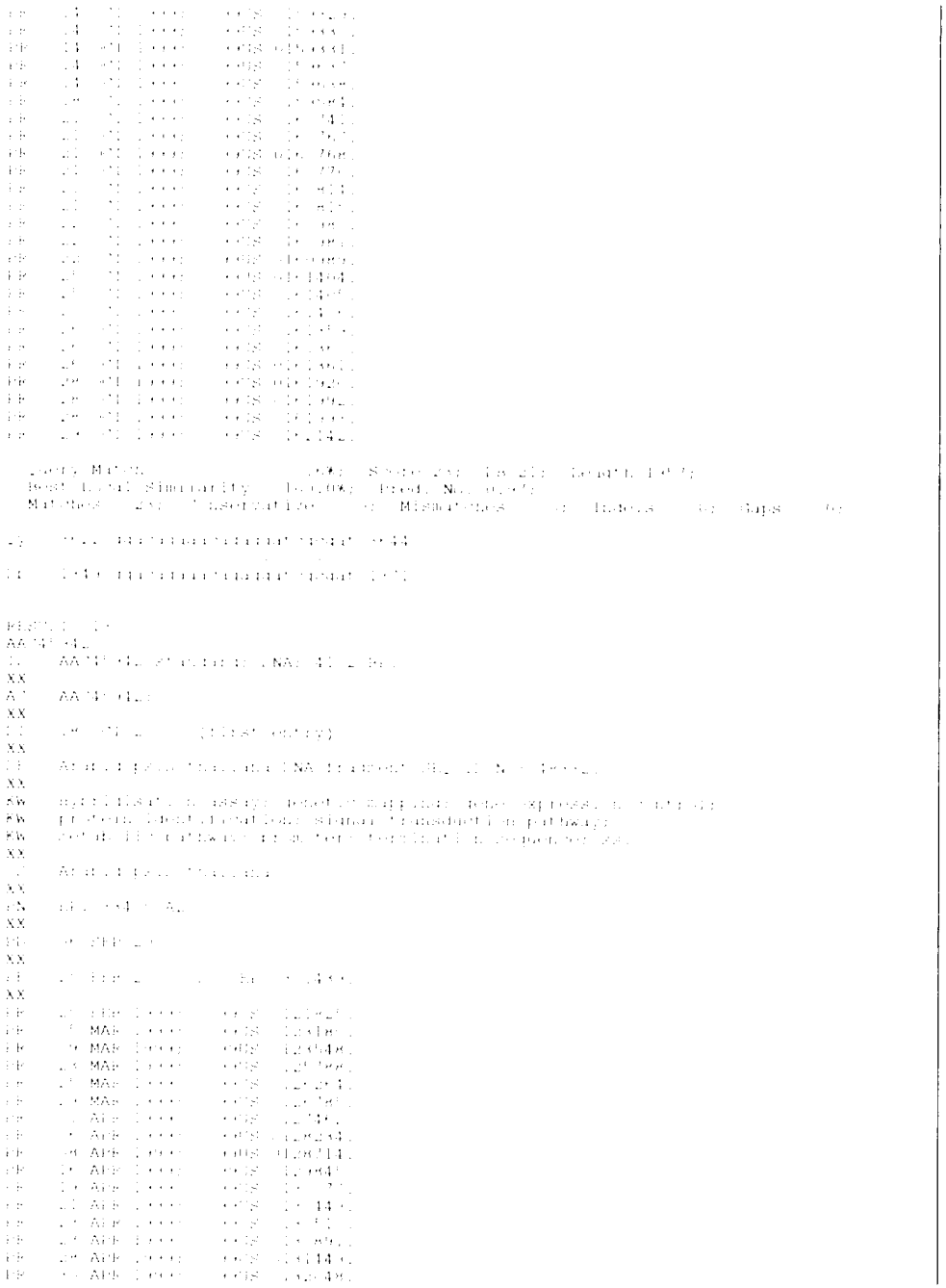


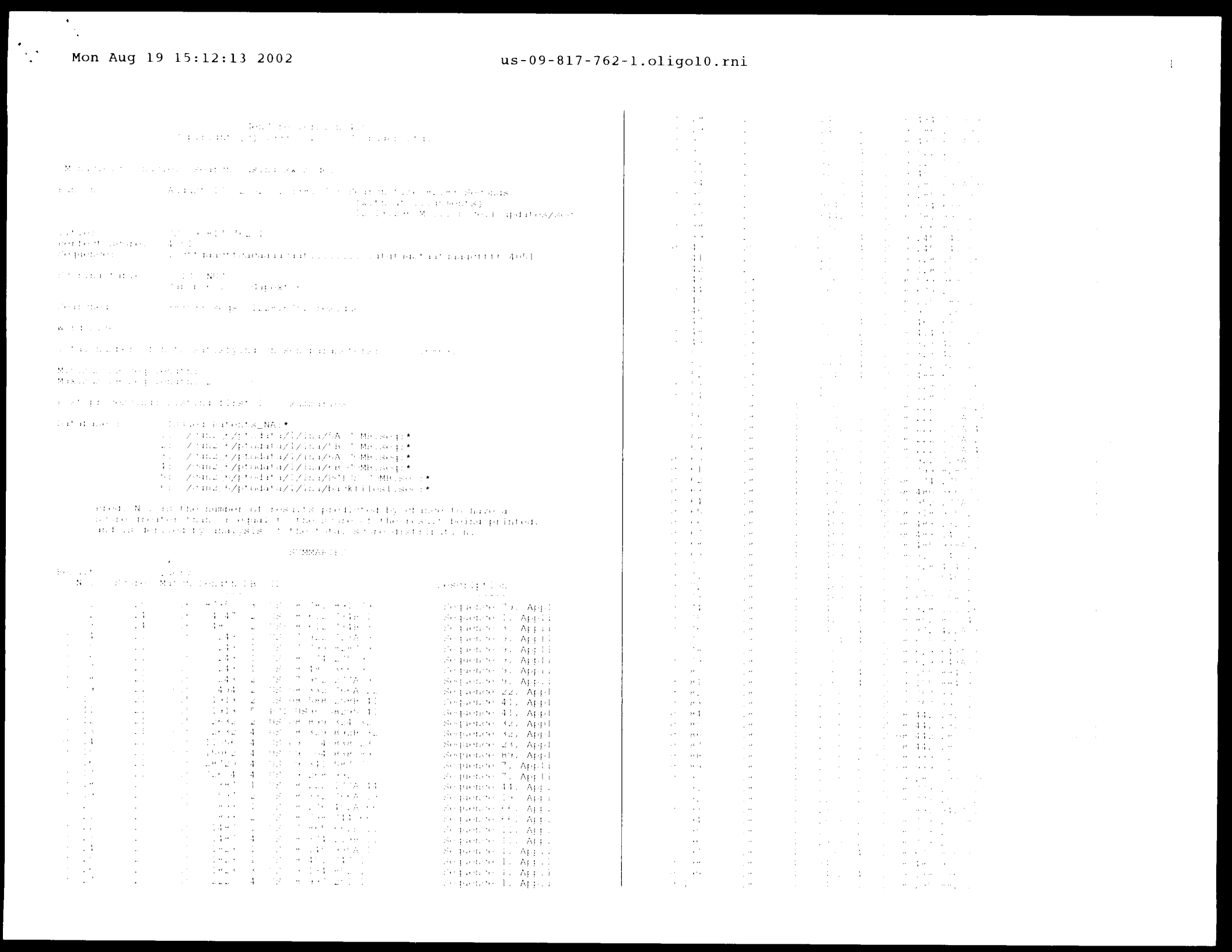
RESULT 2
AAA09020
ID AAA09020 standard; DNA; 14000 BP.
XX
AC AAA09020;
XX
DT 01-AUG-2000 (first entry)
XX
DE A. thaliana AtPAC genomic DNA.
XX
KW Anion channel; AtPAC; mdr homologue; ATP binding cassette transporter;
KW ABC transporter; polycoprotein; inducible; NPP8; herbicide resistance;
KW transgenic plant; ds.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT promoter 1..3429
FT /tag= a
FT /note= "promoter regulatory region"
FT CDS 3429..10428
FT /tag= b
FT exon 3429..3683
FT /tag= c
FT /number= 1
FT intron 3684..3764
FT /tag= d
FT /number= 1
FT exon 3765..3819
FT /tag= e
FT /number= 2
FT intron 3820..4694
FT /tag= f
FT /number= 2
FT exon 4695..4870
FT /tag= g
FT /number= 3
FT intron 4871..5408
FT /tag= h
FT /number= 3
FT exon 5409..5609
FT /tag= i
FT /number= 4
FT intron 5610..5922
FT /tag= j
FT /number= 4
FT exon 5923..6182
FT /tag= k
FT /number= 5
FT intron 6183..6276
FT /tag= l
FT /number= 5
FT exon 6277..6593
FT /tag= m
FT /number= 6
FT intron 6594..6679
FT /tag= n
FT /number= 6

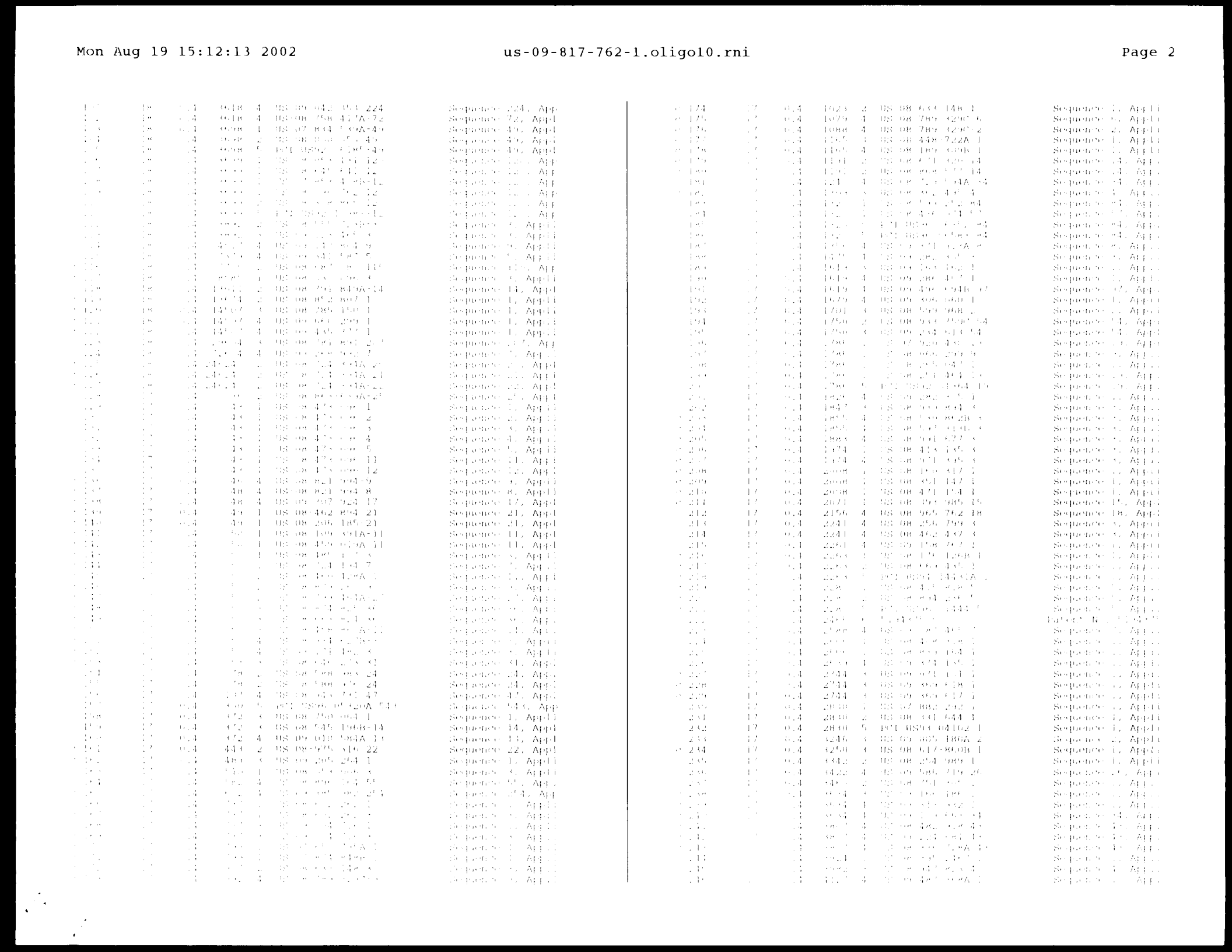


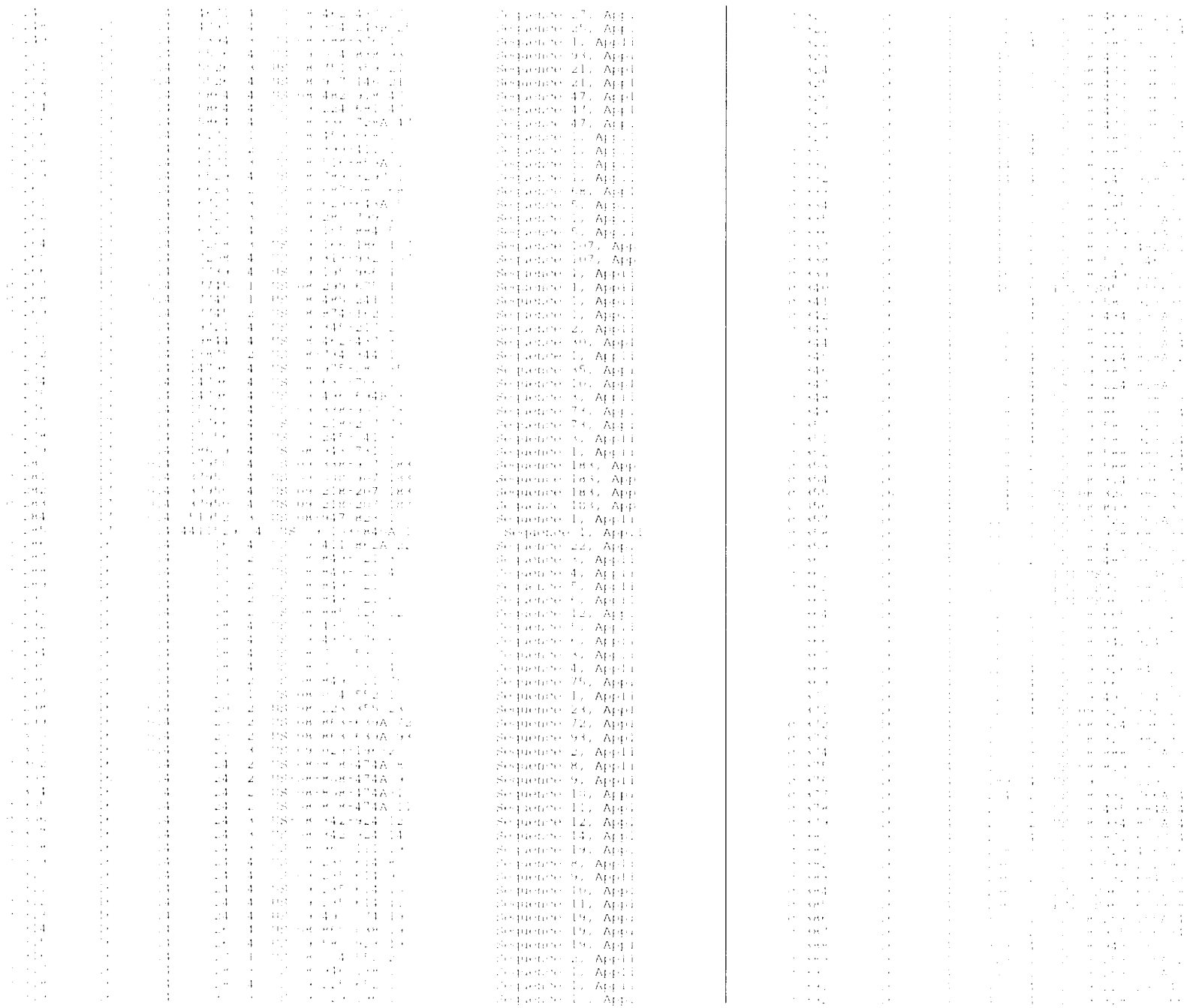


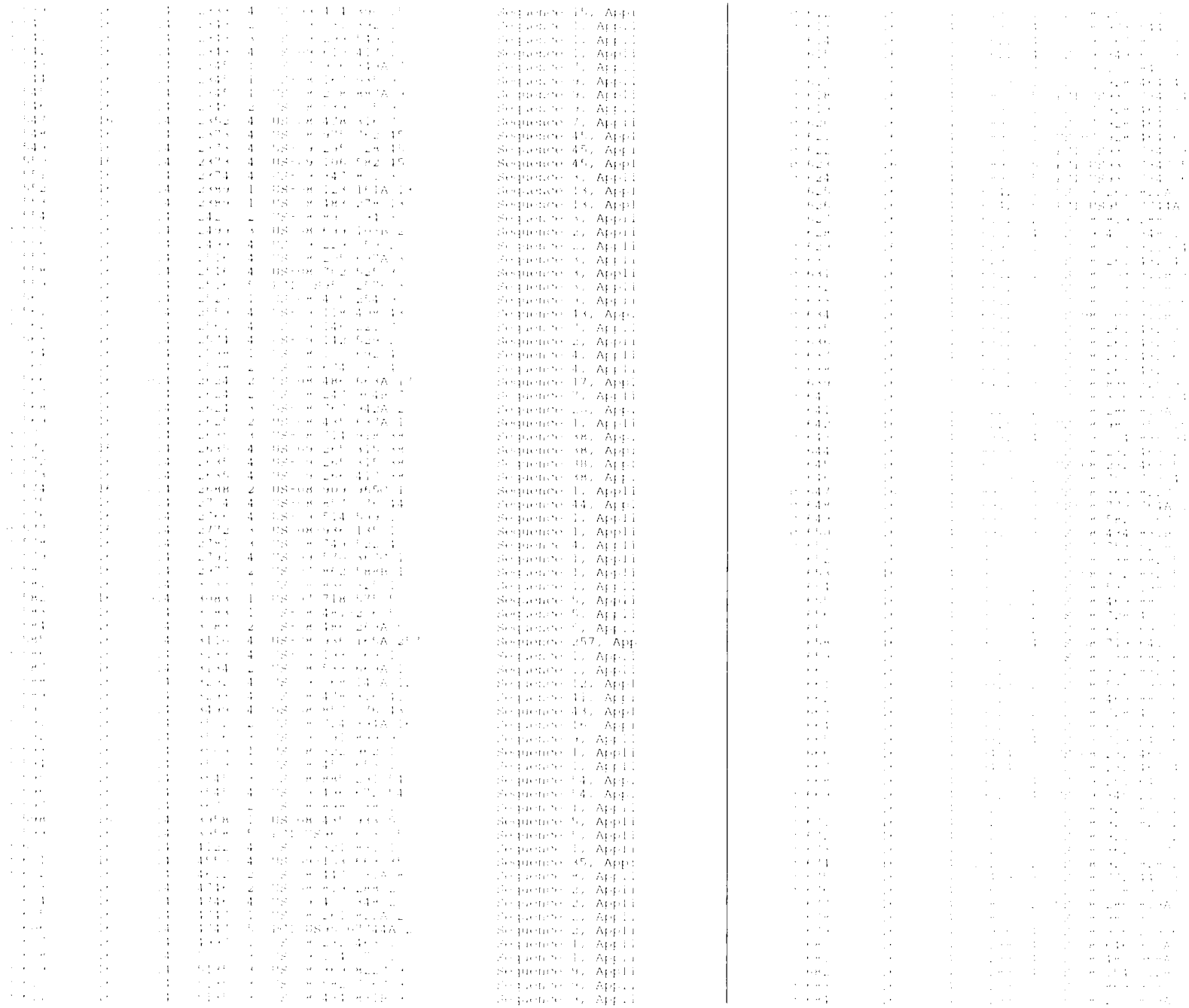




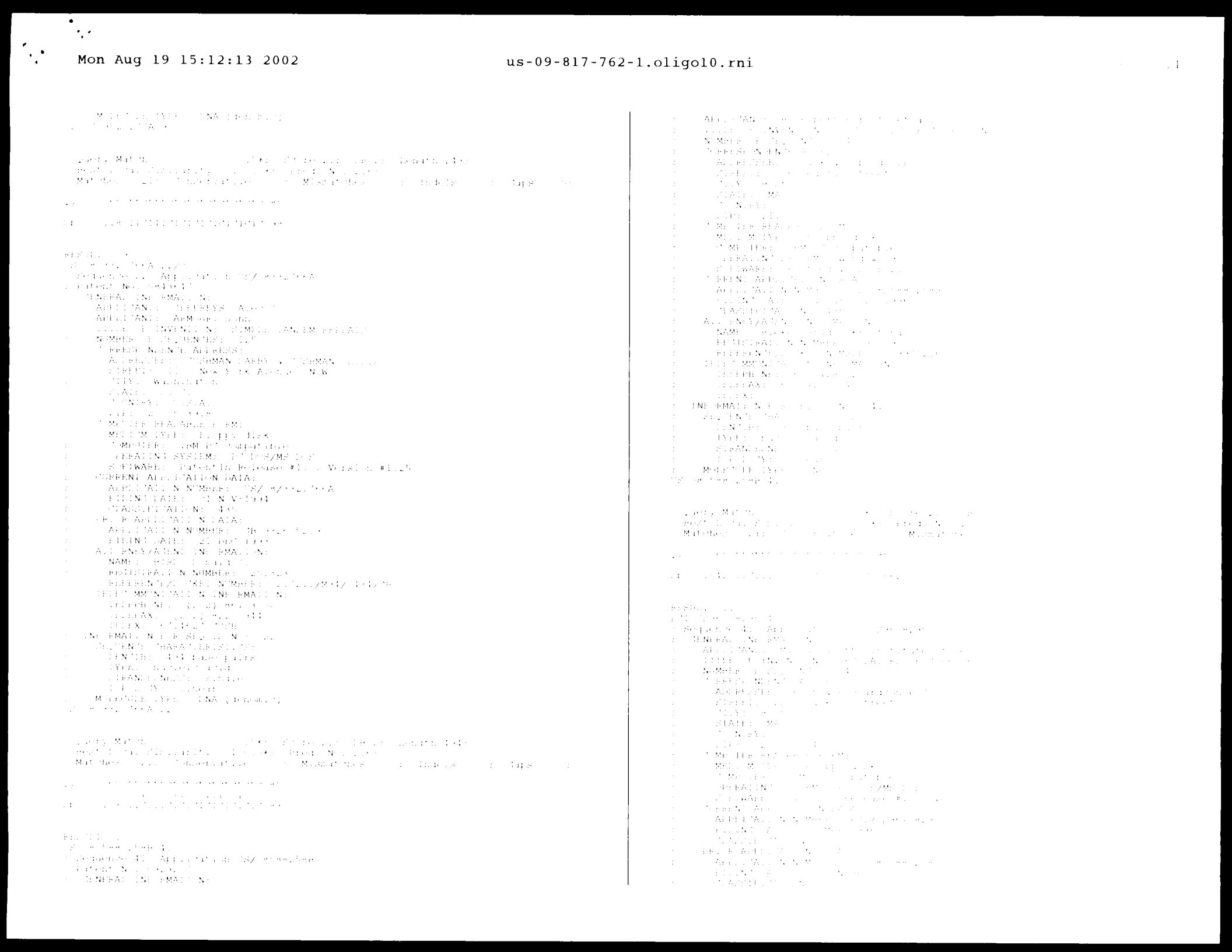






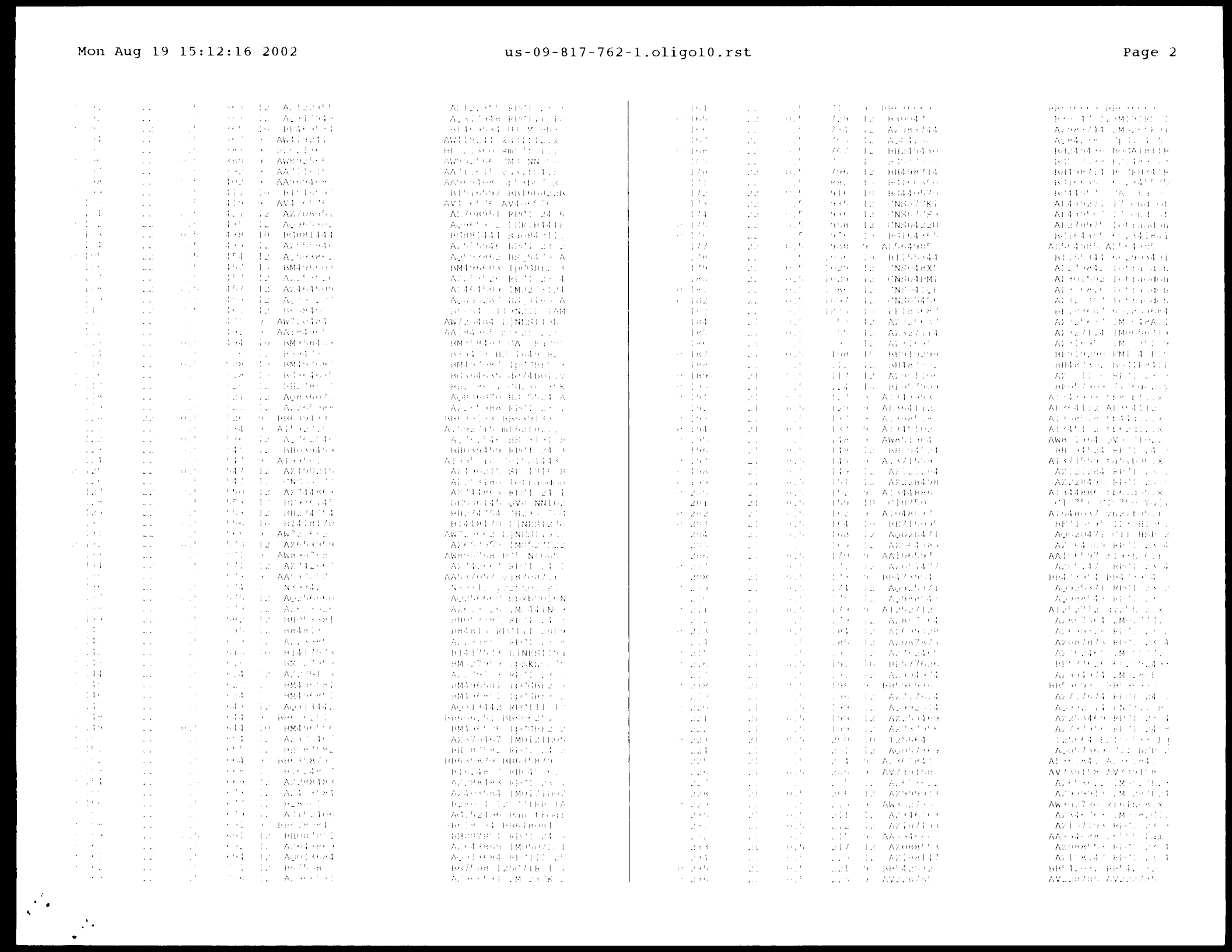


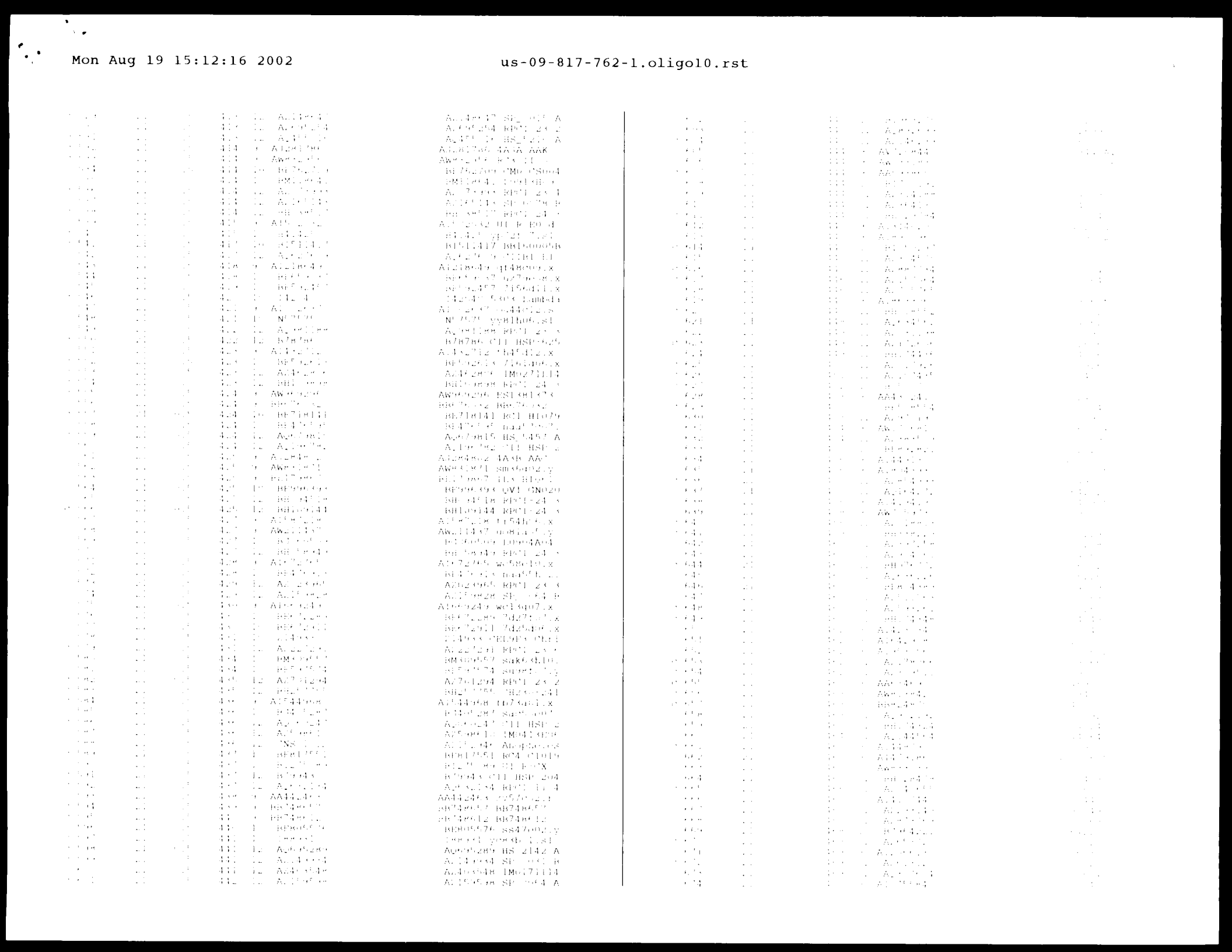


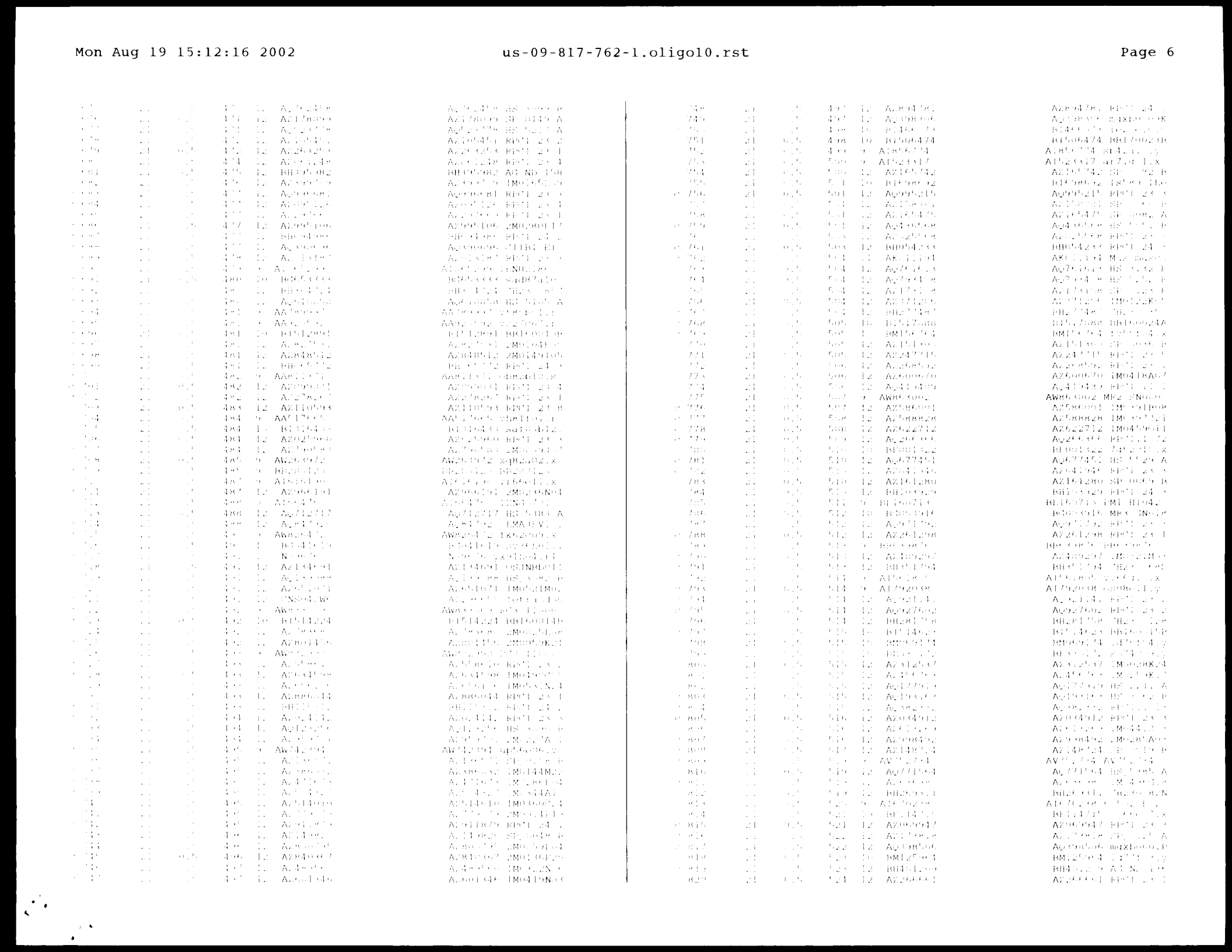


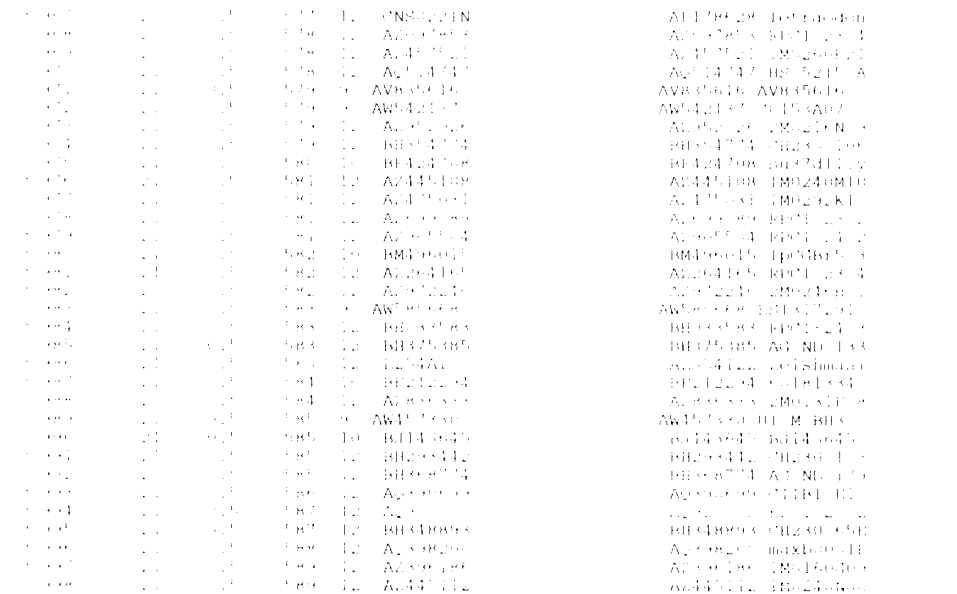
SUMMARIES

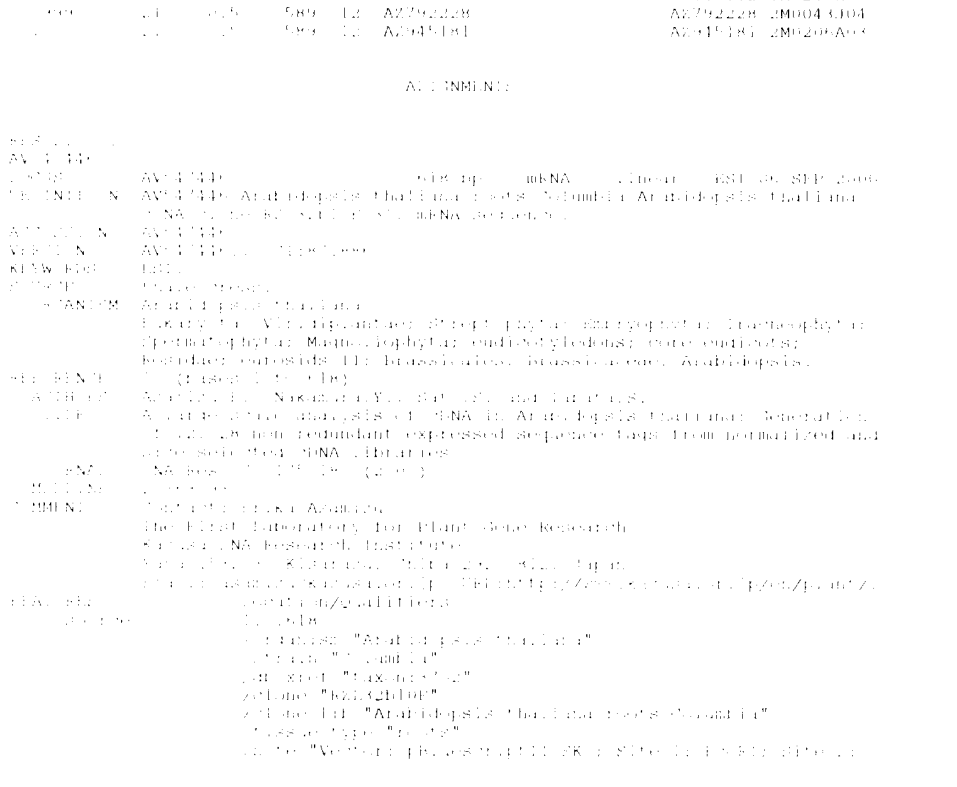
Description

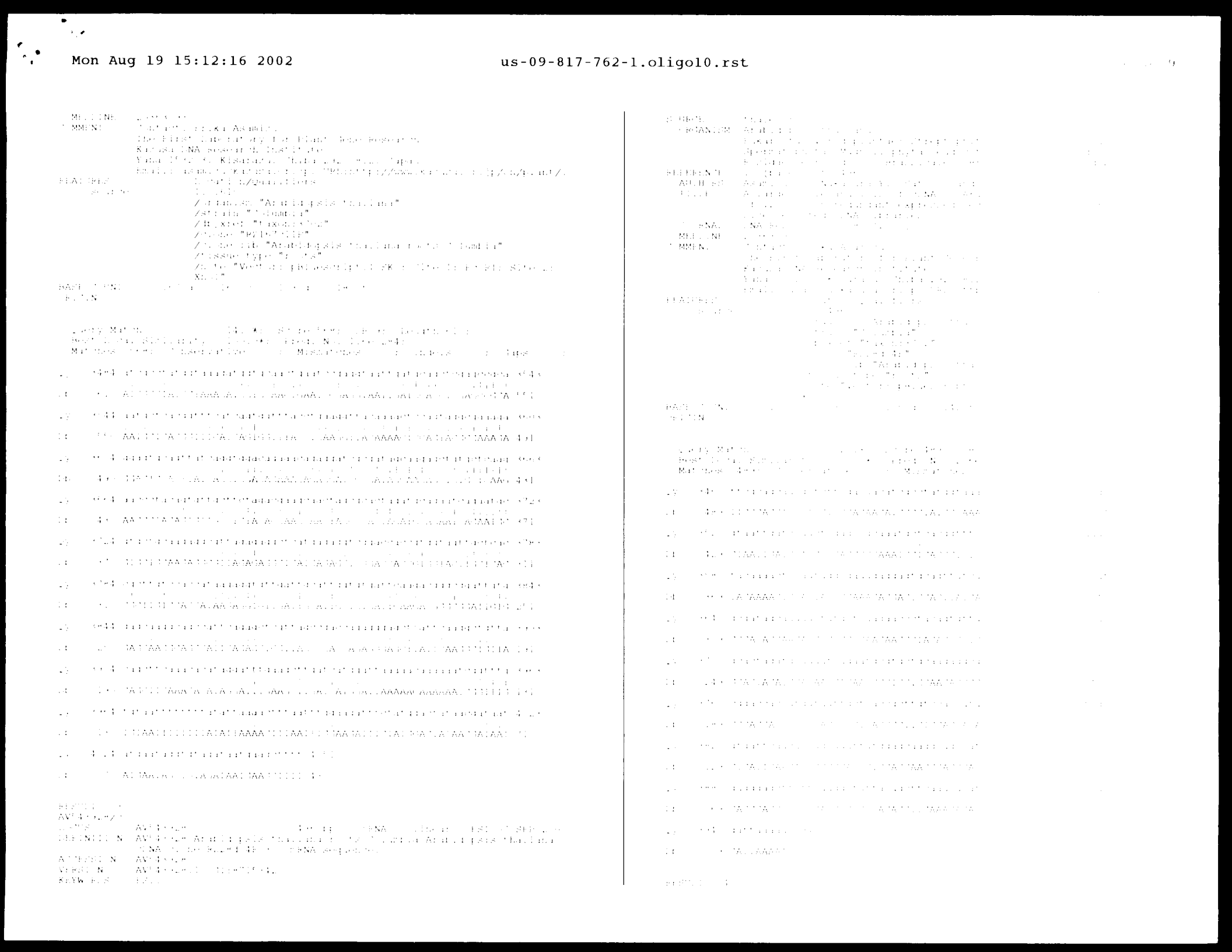


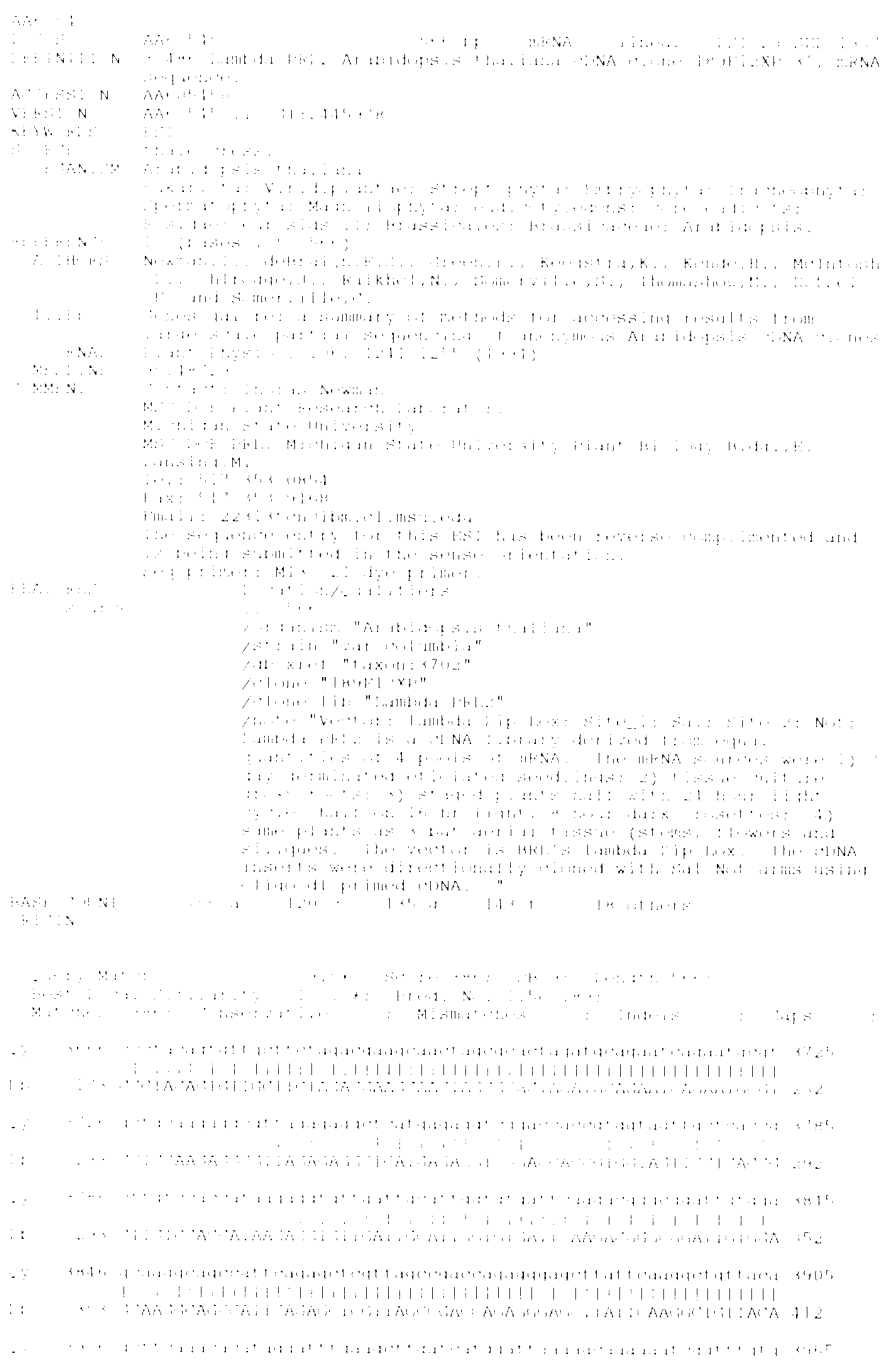


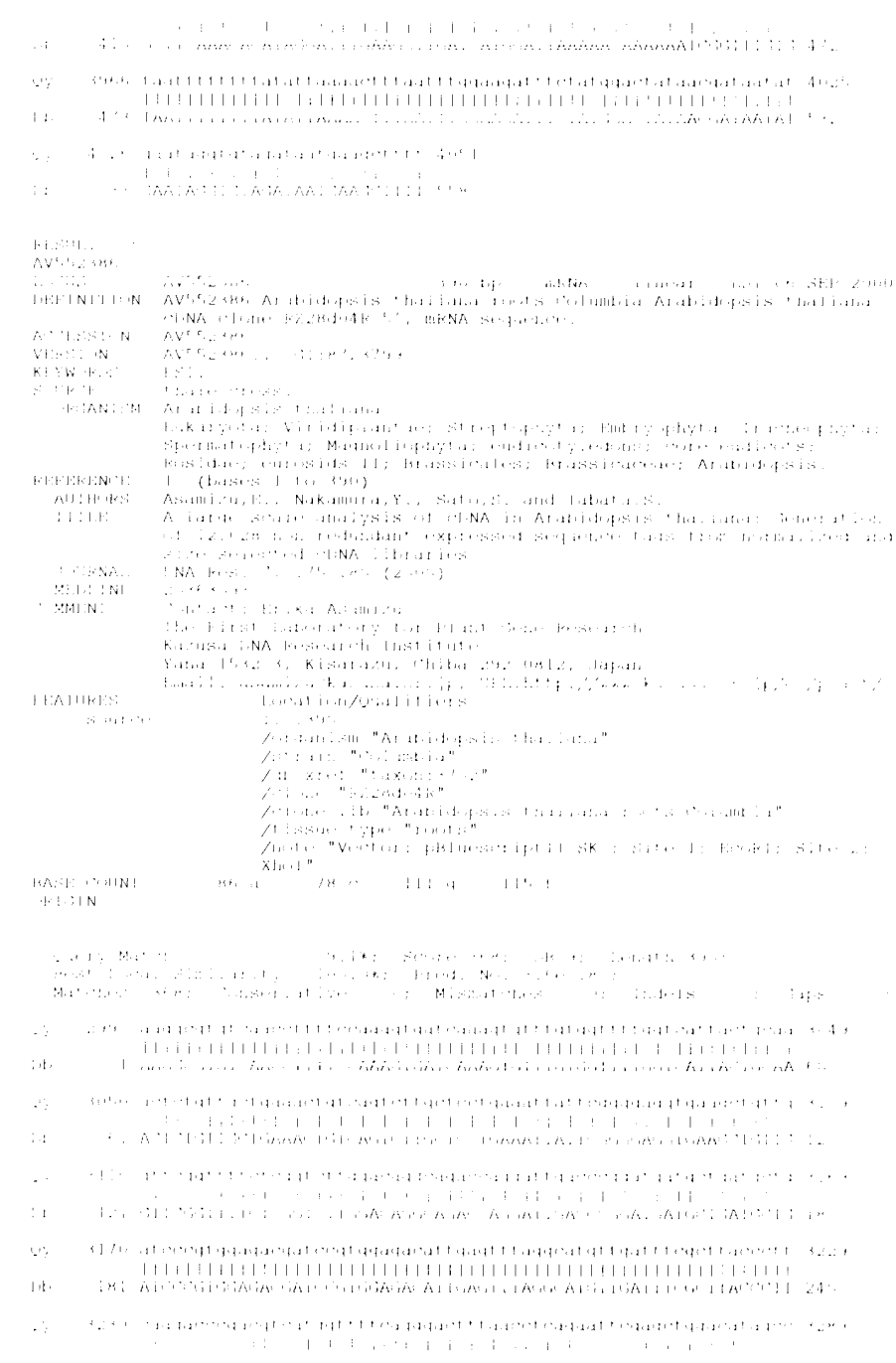


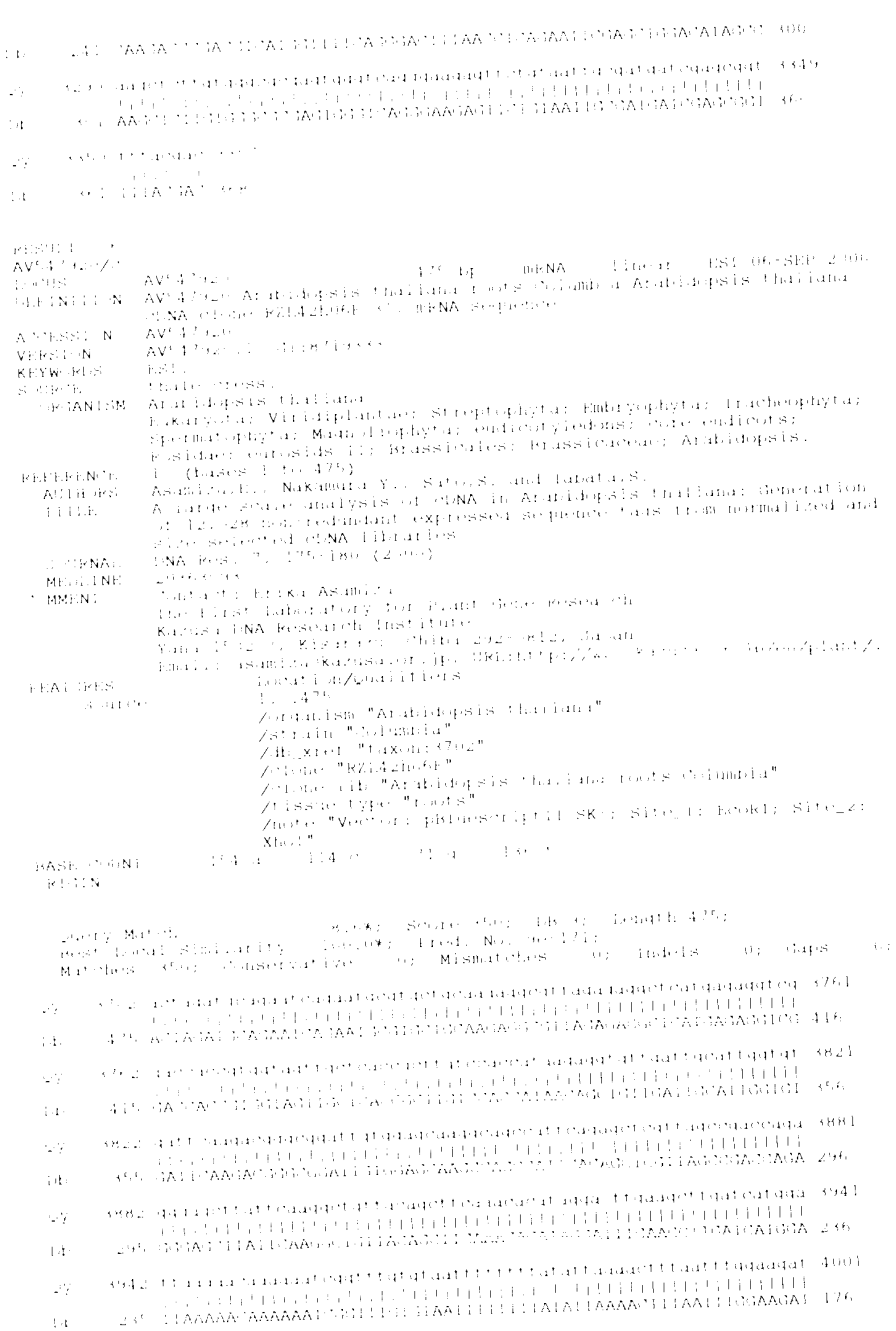


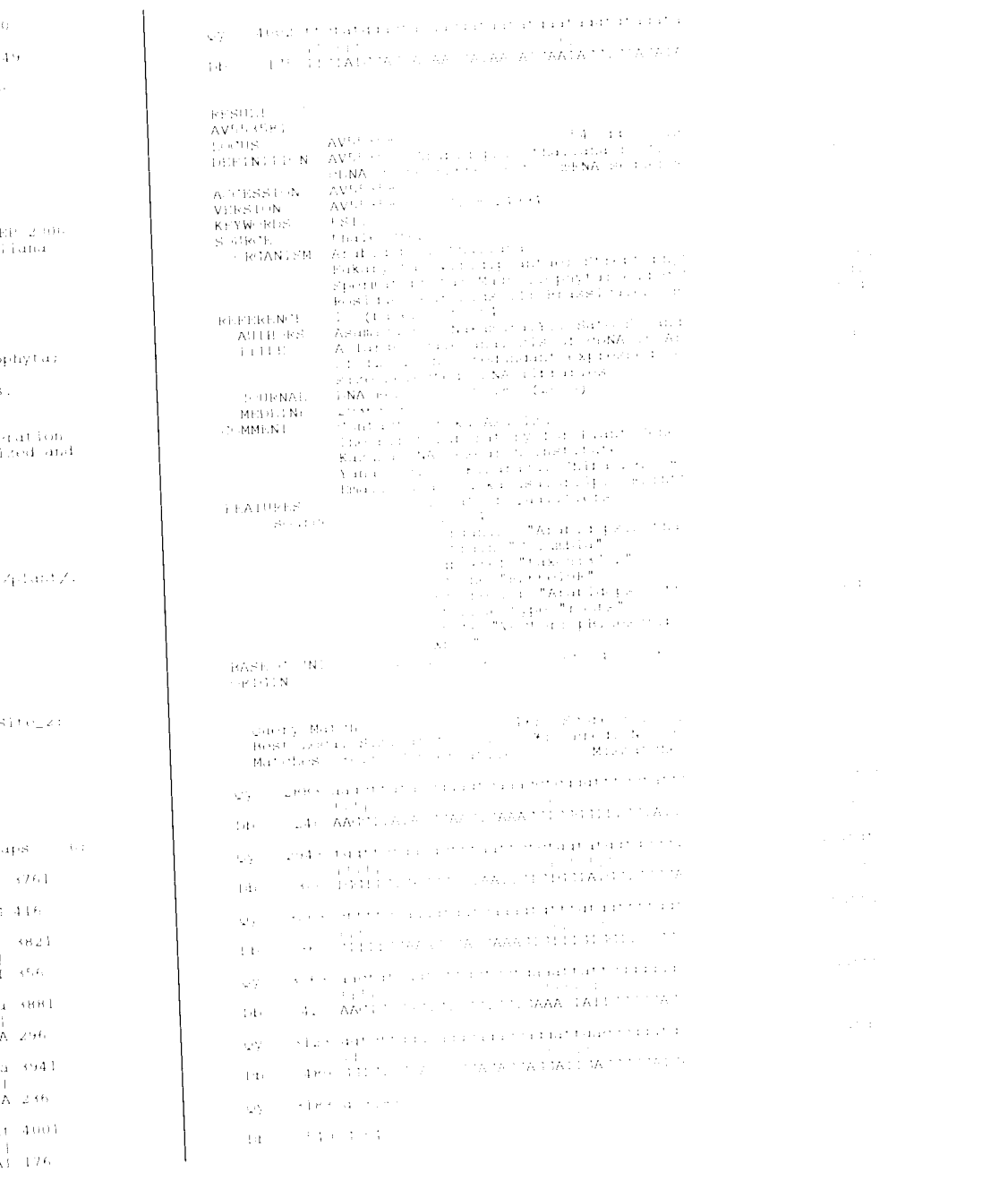












```
RESULT  8
AV964xxx
LOCUS       AV964xxx                 265 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  70150287B A. thaliana, ohio state clone set Arabidopsis thaliana
            cDNA clone 70150287B, mRNA sequence.
ACCESSION   AV964xxx
VERSION     AV964xxx.1  GI:5841238
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Theologis,J., Montiyama,M., Chan,E., Mommey,M., Patroom,B., Gutfilland,D.,
            Wang,X., Hillman,J., Goodlor,K., Kim,S., Boyle,M., Brzoska,P.,
            Gebotag,G., Burns,D., Griffith,J., Monahontona,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Tou,B., Kastury,K., Bonillo,C., Carpio,I.,
            Perlsky,J., Suzuki,C., Argentino,C., Shah,S., Nobrega,A., Murry,L.,
            Turner,S., Kirkorian,S., Elder,L. and Hanson,C.
  TITLE     Arabidopsis thaliana cDNA Differential MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            1354 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877 522 2744
            Fax: 314 423 3424
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1..265
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="70150287B"
                     /clone_lib="A. thaliana, ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana ohio  state clone set."
BASE COUNT        78 a        54 c        66 g        68 t
ORIGIN
```

```
Query Match                 6.2%;  Score 265;  DB 9;  Length 265;
Best Local Similarity   100.0%;  Pred. No. 1.3e-126;
Matches  265;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      154 gatcatgctctatgtaaagctgatgcaacaatattgaagttgaagctgctgctcagct 1599
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GATCATGCTCTATGTAAAGCTGATGCAACAATATTGAAGTTGAAGCTGCTGCTCAGCT 60

Qy      214 gcaaatacgcagctgatttcattgacattgctctcgaagctgacgcagcagcagctgatgaaga 1659
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CGCAATACGCAGCTGATTTCATTGACATTGCTCTCGAAGCTGACGCAGCAGCAGCTGATGAAGA 120

Qy      1660 agtcgtgttgaactctcaagtgaacaaagcagtgaatattgcaattactcaagcgcagtgttg 1719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTCGTGTTGAACTCTCAAGTGAACAAAGCAGTGAATATTGCAATTACTCAAGCGCAGTGTTG 180

Qy      1720 aaaaaccaaagatttcgttatcatgatgaagctaatgcaaagcgtctttatgctagtctatgaa 1779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AAAAACCAAAGATTTCGTTATCATGATGAAGCTAATGCAAAGCGTCTTTATGCTAGTCTATGAA 240

Qy      1780 gatcatgattgaaaaagattgcaaca 1854
             |||||||||||||||||||||||||
Db       241 GATCATGATTGAAAAAGATTGCAACA 265


RESULT  9
AV984xxx
LOCUS       AV984xxx                 353 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV984846 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ154e02F 5', mRNA sequence.
ACCESSION   AV984xxx
```

```
VERSION     AV984846.1  GI:8715848
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/e/index.html
FEATURES             Location/Qualifiers
     source          1..353
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ154e02F"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /tissue_type="roots"
                     /note="Vector: pBluescriptII SK ; Site 1: EcoRI; Site 2:
                     XhoI"
BASE COUNT       129 a       82 c       49 g       93 t
ORIGIN
```

```
Query Match                 6.2%;  Score 250;  DB 9;  Length 353;
Best Local Similarity   100.0%;  Pred. No. 1.3e-118;
Matches  250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     4892 gatattgattgcattgatgtgattgaagcagaagatttgtgaagcaagggcgccattca 4951
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      353 GTGTTGATTGCATTGGTTGTGATTGAAGCAGGAATTGTGGAGCAAGCGCCATTCA 294

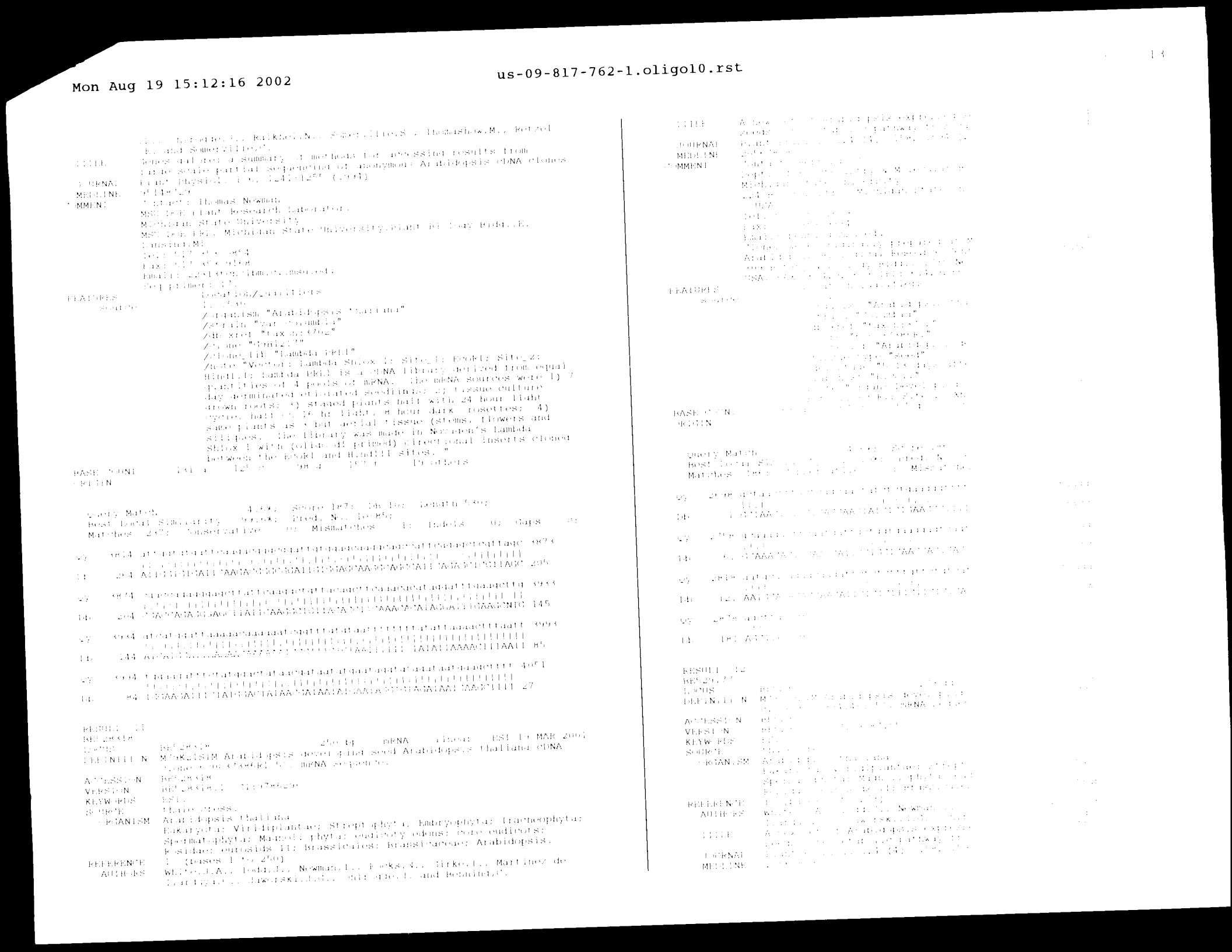
Qy     4952 aagctcgttagcgaacaaagaagaattattcaagcgctattcaagcgctattcaagcattcaaacaatcaag 5023
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      293 CAGCTCGTTAGCGAACAAAGAAGAATTATTCAAGCGCTATTCAAGCGCTATTCAAACAACATCAG 234

Qy     5024 attttaaagctgatgatcatgatgatgaaatcgatttgatgataattttttcg 5081
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      233 ATTTTAAAGCTGATGATCATGATGAAGAAATCGATTTGATGATAATTTTTTCG 180

Qy     5082 aaacttaattgaaaatttgtatgaactatgaatatgatgatgaatgatgaataa 5141
            |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      179 AAACTTAATTGAAAATTTGTATGAACTATGAATATGATGATGAATGATGAATAA 122

Qy     5142 tgaagctttt 5151
            ||||||||||
Db      121 TGAAGCTTTT 112


RESULT  10
N95860/c
LOCUS       N95860                   630 bp    mRNA    linear   EST 05-JAN-1996
DEFINITION  22002 Lambda-ZAP-II Arabidopsis thaliana cDNA clone GOBII-17, mRNA
            sequence.
ACCESSION   N95860
VERSION     N95860.1  GI:1265179
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L.
```



COMMENT Author(s): Weigel, D.
Dept. of Biochemistry & Molecular Biology
Michigan State University
Dept. of Chemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 9724
Fax: 517 353 1794
Email: weigel@msu.edu
Publication: Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dooner HK, Pereira A, Nam HG
Activation tagging in Arabidopsis.
Plant Physiol 122(4):1003-13 (2000)

PUBMED 10759499

COMMENT Author(s): Weigel, D.
Dept. of Biochemistry & Molecular Biology
Michigan State University
East Lansing, MI 48824
USA
Tel: 517 355 9724
Fax: 517 353 1794
Email: weigel@msu.edu

Publication: Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dooner HK, Pereira A, Nam HG
Activation tagging in Arabidopsis.
Plant Physiol 122(4):1003-13 (2000)

FEATURES Location/Qualifiers
source 1..?
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="lib2xAI"
/clone_lib="lambda ZAP2"
/note="Vector: lambda Zap Express; Site_1: Sal1; Site_2: Not1;
Lambda ZAP2 is a cDNA library derived from equal
contributions of 4 pools of mRNA: 1) mRNA from
dry dormant/germinated seeds(dies); 2) tissue culture
grown rosettes; 3) stunted plants left with 24 hour light
cycles, half an LP or IRRI, 8 h a dark treatment; 4)
some plants as 3 but petiole tissue (stems, flowers, and
siliques. The pool of 20 mRNAs lambda zap cDNA. The cDNA
inserts were directionally cloned with Sal No1 ends ligated
to oligo dT primed cDNA "

PUBMED 10759499

Query Match 1.2%; Score 149; DB 4; Length ? ;
Best Local Similarity 100.0%; Pred. No. 3.6e-07;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ...
Db 1 ...

Qy 1 ...
Db 1 ...

Qy 1 ...
Db 1 ...

Qy 1 ...
Db 1 ...

RESULT 13
PE5_seq:
LOCUS BE5_seq 207 bp mRNA linear EST 27-MAR-...
DEFINITION M70A06T1M Arabidopsis thaliana seed Arabidopsis thaliana cDNA
clone ..., mRNA sequence.
ACCESSION BE5_seq
VERSION BE5_seq.1 GI:7...
KEYWORDS EST.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to ...)
AUTHORS Whitelaw,A., Todd,J., Newman,T., Parks,N., Kirkpatrick, Martinez &
Hartzenberg,... Lewenski,K.C., Editor(s), and Bernatzky,...
TITLE A new set of Arabidopsis expressed sequence tags from developmental
seeds: The potential for transposon tagging and functional
...
JOURNAL Plant physiol. 124 (4), 1596 ... (2000)
MEDLINE ...
COMMENT Author(s): Weigel, D.
Dept. of Biochemistry & Molecular Biology
Michigan State University
...Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 ... 9724
Fax: 517 ... 1794
Email: weigel@msu.edu

